Characterisation of KRAB-ZFPs in the mouse

Stephanie Briers

james s. mcdonnell foundation

MRC Medical Research Council
Human Genetics Unit

Thesis presented for the degree of Doctor of Philosophy,
University of Edinburgh
~2005~

Declaration

The experiments described in this thesis are the unaided work of the author except where acknowledgement is made by reference. No part of this work has been previously accepted for any other degree or professional qualification, nor is it in the process of being submitted for any other degree.

Stephanie Briers
January 2004

"Where then does wisdom come from?

Where does understanding dwell?

It is hidden from the eyes of every living thing……,

God understands the way to it and he alone knows where it dwells,

……And he said to man,

'The fear of the Lord-that is wisdom, and to shun evil is understanding.' "

Job 28: 20-28, The Bible

Acknowledgements

Woohoooo! At last, I'm writing my acknowledgements! This might take a while.

Firstly, in my quest for a PhD project, none really hit the spot until I met Wendy and was introduced to the world of nuclear organisation and KRAB-ZFPs. I didn't realise back then the caliber of the group I was joining, or how dedicated a group leader you were. Your concern about other peoples understanding of science (students and public alike) has been inspirational to me and I will never be able to thank you enough for giving me a shot at this PhD and seeing me as a prospective research scientist. Although my culinary achievements were indeed like some of my preliminary science experiments - hit and miss, I'm glad to say that both have vastly improved whilst being a member of your lab. Both of these are also attributed to your team of post-docs and technicians- if not for them, my time in Edinburgh would have been far less instructive and less memorable.

On these lines my thanks goes to all of yous lot in the Bickmore lab, both past and present members. Firstly, thanks goes to the students (Nic the girl, Inga Winger, Anne the Pan, Gill Ring (surrogate member) and, of course, Sue Gilchrist) for making the lab a fun place to work. I want to say a special thanks to the latter two for their good advice and lending me their ears through from time to time. Oh and I mustn't forget their vital tuition of the Glasweegen (sp?) accent, without which I might as well have done my PhD in France with Sev, ha ha ha! Speaking of mad, Shelagh (the 2nd maddest person in the lab) also deserves recognition for tuition of all things Scottish, FISH advice, being a great hostess and of course stupendous cakes (I hope that word means what I think it does?). To NTB my dancing buddie, thanks for eating my 'mare' of a cheese fondue and for pointing out to me the somewhat frequent occasions of when I was being a nerd, or as you more tenderly put 'f^%*#g stupid'. Always brings a smile to my face when I think of that one☺. Thanks also to Sev, (aka. the mad French bird). I will miss being in an office with you. We've had some good laughs! Remind me never to ask for your help in making caramel sauce though. Hmmm, maybe you just didn't understand what I was saying?

To Mary T, thanks for being such good company and an excellent problem solver of both work and wedding related dilemmas. Thanks also for your generous provisions ranging from various coffee-break treats, cup o' soups and even breakfast on the odd occasion! To PP and Sandy, cheers both of you for everything. From fixing door locks to computer aid, I'm truly grateful. Remember the veggie oil and you heard it here first! A special thanks to Gogo, Kev and co., and to all the animal house staff for your invaluable contribution towards this thesis.

In saving the best till last, I would like to say a special and sincere thank you to Heidi. In my new post-doc position, I will strive to be as good a post-doc as you. From your selflessness, patience and diplomacy with everyone in the lab to your continuous encouragement, support and friendship, you have given me the

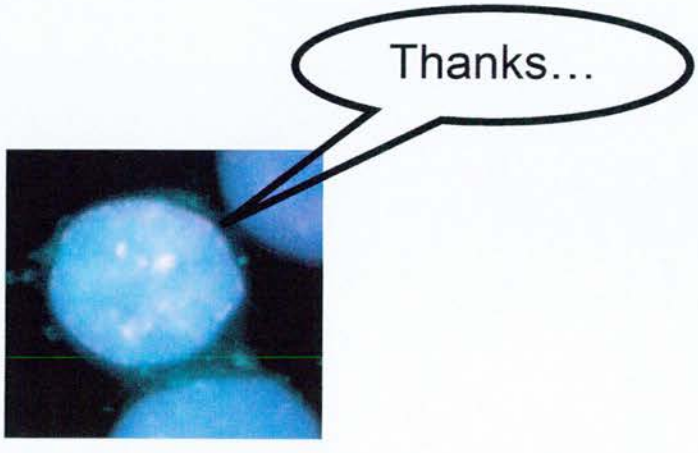
confidence to not give up and keep pushing for those results. Yep things didn't always go the way we thought, but working with you has been a pleasure and I would not trade this experience for anything.

To my best friends, Shiv and Gill, who witnessed the tears, and picked me up in ways that only you two can. I love you both and I'll be seeing you soon xx. Thanks also to Catherine, Charlotte, Amy, Laura and Nic for all your encouraging emails, cards and surprise presents, it was good to know you were there.

In closing, I am of course indebted to all my friends and family (old and new members) and always will be for their unwavering source of love, support and prayers and I look forward to spending more time with you all soon. To my nan, grandad, and auntie Cal, thanks especially for being patient with me when I've snapped. To our Paul, Matt and Joe, thanks for making me laugh and taking my mind off stuff when I came home. I've missed you loads.

To my dad, you are still the wisest person I've ever met. You used to say that I' d never be as smart as you, and you are still right. There's nothing to say except thank you and I love you.

And finally, the very last person to thank is my husband. Three years ago the thought of doing a PhD seemed like a dream come true. The thought of getting married to you as well was just TOO good to be true. I've been blessed with both and I thank God. Love you loads. x.



Abstract

KRAB-ZFPs (Krüppel associated box-zinc finger proteins) are known to be involved in the transcriptional repression of genes. Direct interaction of the KRAB domain with the co-repressor, KAP-1, recruits other heterochromatin-associated proteins such as HP1 and histone deacetylases to the DNA and results in gene silencing during vertebrate development. Although KRAB-ZFPs comprise the largest transcription factor family in mammals, little is known of their specific biological functions and functional redundancy has been postulated. Moreover, few KRAB-ZFPs have known target genes.

In 2001, Sutherland *et al.*, performed a gene-trap screen in mouse ES cells for nuclear proteins. Seven novel KRAB-ZFPs were identified. I investigated the sub-nuclear distribution of the trapped fusion proteins and found that all except one co-localised with KAP-1 at pericentromeric heterochromatin in a subpopulation of cells. Furthermore this sub-population of cells increased when induced to differentiate with retinoic acid. I further investigated the localisation of these fusion proteins by making full-length GFP-tagged constructs with KRAB-ZFP 492, and by producing an antibody to the endogenous protein. KRAB-ZFP 492 mis-localised in the gene-trap screen as AP492 antibody identifies both nuclear diffuse and nucleolar staining in ES cells. This suggests that the zinc finger domain of KRAB-ZFP 492 is required for correct localisation of this protein *in vivo*. Furthermore, endogenous KRAB-ZFP 492 is not recruited to pericentromeric heterochromatin via KAP-1 upon differentiation. However in differentiated cells, endogenous KRAB-ZFP 492 and NT-2 is found associated with KAP-1 and HP1 proteins in smaller foci. This suggests that KRAB-ZFP target genes may not be recruited to heterochromatin for repression.

To gain insight into the functions of KRAB-ZFPs, I generated two mutant KRAB-ZFP mouse lines using gene-trap ES cells. From these mice I was able to detect the tissues in which each KRAB-ZFP was expressed. In $113^{+/-}$ embryos, fusion protein expression is initially restricted to the heart and thymus. In adults, expression is

observed in the thymus. Expression is less specific in $492^{+/-}$ embryos. KRAB-ZFPs often lie in clusters of very similar genes, hence the redundancy of KRAB-ZFPs 113 and 492 are being tested by examining the phenotype of homozygous mutants for each gene. Information gained through these studies and the reagents generated will enable the target genes for these proteins to be determined in the future.

Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Abbreviations

2-D	two-dimensional
3-D	three-dimensional
A	adenine
Ac	acetate
ActD	Actinomycin D
ala	alanine
ADP	adenosine diphosphate
ATP	adenosine triphosphate
AP	affinity purified
BAC	bacterial artificial chromosomes
bp	base pairs of DNA
BRCA-1	breast cancer associated gene-1
BrdU	bromodeoxyuridine
BSA	bovine serum albumin
C	cytosine
CBP	CREB-binding protein
cDNA	complementary deoxyribonucleic acid
CENP	centromere protein
chromodomain	chromatin organiser modifier
Ci	Curies
cys	cystine
CpG	cytosine and guanine
CREST	calcinosis, Rayaud's phenomenon esophageal dismotility, sclerodactyly, telangiectasia
C-terminal	carboxy-terminal
Da	Daltons
DAPI	4,6-diamine-2-phenylindole
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanine triphosphate
dH_2O	distilled water
DIC	differential interference contrast
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTPs	deoxynucleotide triphosphates
dpc	days post coitum
Drosophila	*Drosophila melanogaster*
ds	double strand

DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
E	embryonic day
E. coli	*Escherichia coli*
EDTA	ethylenediaminotetra-acetic acid
EGTA	ethylene glycol bis (β-aminoethyl ether) N, N, N', N'-tetraacetic acid
ER	endoplasmic reticulum
ES	embryonic stem
EtBr	2,7-diamino-10-ethyl-9-phenyl-phenathridium (ethidium bromide)
EtOH	ethanol
Fab	fragment antibody binding
FCS	foetal calf serum
FISH	fluorescence *in situ* hybridisation
FITC	fluorescein isothiocyanate
g	gravities
G	guanine
G1	growth phase 1 of the cell cycle (pre-replication)
G2	growth phase 2 of the cell cycle (post-replication)
GFP	green fluorescent protein
gln	glutamine
gly	glycine
GMEM	Glasgow's modified Eagle's medium
GST	Glutathione-S-transferase
HA	haemagglutinin
HAT	histone acetyltransferase
HDAC	histone deacetylase
his	histidine
HMT	histone methyltransferase
HP1	heterochromatin protein 1
hr	hours
HRP	horseraddish peroxidase
HSA	human autosome
ICF	Immunodeficiency, Centromeric instability, and Facial syndrome
IF	immunofluorescence
IgG	immunoglobulin G
IPTG	Isopropyl bD-thiogalactopyranoside
K	lysine
Kb	kilobase pairs of DNA
kDa	kilo Daltons (molecular weight/10^3)

KLH	keyhole limpet hemocyanin
k/o	knockout
KRAB-ZFP	Krüppel associated box -zinc finger protein
LB	Luria-Bertani
leu	leucine
LIF	leukaemia inhibitory factor
lys	lysine
M	molar
MAA	methanol/acetic acid
mAb	monoclonal antibody
Mb	megabase pairs of DNA
MBD	methyl binding domain
β-ME	2-β-mercaptoethanol
MeCP	methyl binding protein
met	methionine
Met_3	trimethylation
min	minutes
MMU	mouse autosome
MOPS	3- (N-Morpholino) propanesulfonic acid; 4-Morpholinepropanesulphonic acid, pH 7
mRNA	messenger RNA
Mw	molecular weight
nfr	nuclear fast red
NLS	nuclear localisation signal
NP-40	Nonidet P-40
NPD	nuclear protein database
N-terminal	amino-terminal
NTP	nucleotide triphosphate
OD	optical density
oligos	oligonucleotides
o/n	overnight
ORF	open reading frame
p	short arm of chromosome
pg	passage
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PcG	polycomb group
PCR	polymerase chain reaction
PEV	position effect variegation
pFa	paraformaldehyde
pI	Isoelectric point
PML	promyelotic leukaemia

PMSF	phenyl methyl sulfonyl fluoride
pro	proline
PVDF	polyvinylidene difluoride
q	long arm of chromosome
R	rhodamine
R-band	chromosome pattern produced by reverse Giemsa staining
RNA	ribonucleic acid
RNAse	ribonuclease
RNO	rat autosome
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	room temperature
S. cerevisiae	*Saccharomyces cerevisiae*
sec	seconds
ser	serine
SDS	sodium dodecyl sulfate
ss	single strand
SSC	standard sodium citrate
SPF	specific pathogen free environment
S-phase	DNA synthesis phase of the cell-cycle
S. pombe	*Schizosaccharomyces pombe*
SWI/SNF	mating type switching/sucrose non-fermenting
T	thymine
TAE	Tris-EDTA acetic acid buffer, pH 8
TBS	Tris-buffered saline
TE	Tris-EDTA, pH 8
TEMED	N, N, N', N-tetramethylethylene diamine
TdT	terminal deoxynucleotidyl transferase
thr	threonine
TPS	Total protein sample
TR	Texas Red
TSA	Trichostatin-A
U	Unit
UTR	untranslated region
UV	ultraviolet
v.	Version
vol	volume
v/v	volume/volume
wt	wild type
w/v	weight/volume

X-Gal	5-Bromo-4-Chloro-3-indolyl-αD-galactoside
Xi	inactive X chromosome
Xenopus	*Xenopus laevis*
ZF	zinc finger

Chapter 1

Introduction

Our linear DNA sequence consists of ~6.6x 10^9 bp of DNA, which when fully extended measures ~2 m in length. Yet, it can be packaged into a nucleus barely 15 μm in diameter (i.e. one hundred thousand times smaller) by its hierarchical organisation into chromatin. The DNA double helix alone, although fundamental to our chromosomal inheritance, is no longer considered to be definitive to our genetic make-up. Rather the rules that govern the epigenetic regulation of expression of our ~24,000 protein-coding genes (Ensembl release 27.35a.1, December 2004) must lie at the level of chromatin structure.

Proteins with related functions in the cytoplasm are often concentrated together in three-dimensional (3-D) space, within the confines of discrete membrane bound organelles, such as mitochondria and the Golgi apparatus. Similarly, the mammalian nucleus is organised into domains associated with different facets of nuclear function, despite the absence of membrane bound compartments. Nuclear compartments are physical spaces where specific proteins, or protein complexes, are concentrated even though these proteins are in dynamic flux with the rest of the nucleoplasm (Chubb and Bickmore, 2003). It is now clear that the confinement of biomolecules such as DNA and proteins within such nuclear compartments is crucial for correct cell function, and disruption of this often results in human genetic disease, cancers and during viral infection (Marsh *et al.*, 1998; Koken *et al.*, 1997; Bell *et al.*, 2000; Chen *et al.*, 2004).

1.1 Gene regulation by *trans*-acting factors and *cis*-acting DNA elements

The transcriptional status of every gene is controlled by at least two components of a genetic switch; regulatory stretches of DNA *in cis*, such as the promoter and enhancer sequences, and transcription factors that promote gene activation or repression (Alberts *et al.*, 1994). Genetic analyses in bacteria provided the first evidence for DNA-binding transcription factors, the λ repressor and the *lac* repressor (Ptashne, 1967; Muller-Hill *et al.*, 1968). Prokaryotic repressors such as the λ

repressor work in the simplest case by steric hindrance (Fig. 1.1A). Positive regulation of prokaryotic genes occurs by either removal of a repressor protein from DNA, or by binding of an activator protein (Fig. 1.1A). The *lac operon* is an example of a more complex genetic switch where positive and negative controls combine (Muller-Hill *et al.*, 1968).

The same basic strategies as above are used to control gene expression in eukaryotes, but the genetic switches that are used are much more complex. Unlike prokaryotic repression, transcriptional repression in eukaryotes is not dictated by the direct competition of transcriptional repressors and RNA polymerases for access to DNA (Fig. 1.1B). Firstly, eukaryotic RNA polymerases cannot initiate transcription on their own. They require a set of 'basal transcription factors', which must be assembled at the promoter before transcription can start. The assembly of the 'transcription machinery' is required for initiation by RNA polymerase II (Pol II). A second difference between prokaryotic and eukaryotic transcription factors lies in the location of DNA sequences they are able to bind. In eukaryotes, many gene regulatory proteins can bind to DNA elements, for example, enhancer and/or insulator sequences, thousands of nucleotides away from the promoter they influence (reviewed by Cook, 2003; Sage *et al.*, 2005). In many cases the DNA between these regions and the promoter loops out to allow the bound transcriptional regulator to physically interact with RNA polymerase II or other members of the transcription machinery. Sophisticated experimental techniques such as 'chromosome conformation capture' and RNA TRAP (tagging and recovery of associated proteins) have provided clear evidence of looping at the β-globin locus (Tolhuis *et al.*, 2002). The effect of looping out enables proteins tethered at the distant region to collide repeatedly with proteins bound at the promoter, which has the same effect as would be obtained by increasing the protein's local concentration at the promoter. In order to study the mechanisms of gene regulation by transcription factors it is essential to understand how their structures enable interaction with DNA and/or RNA and other proteins.

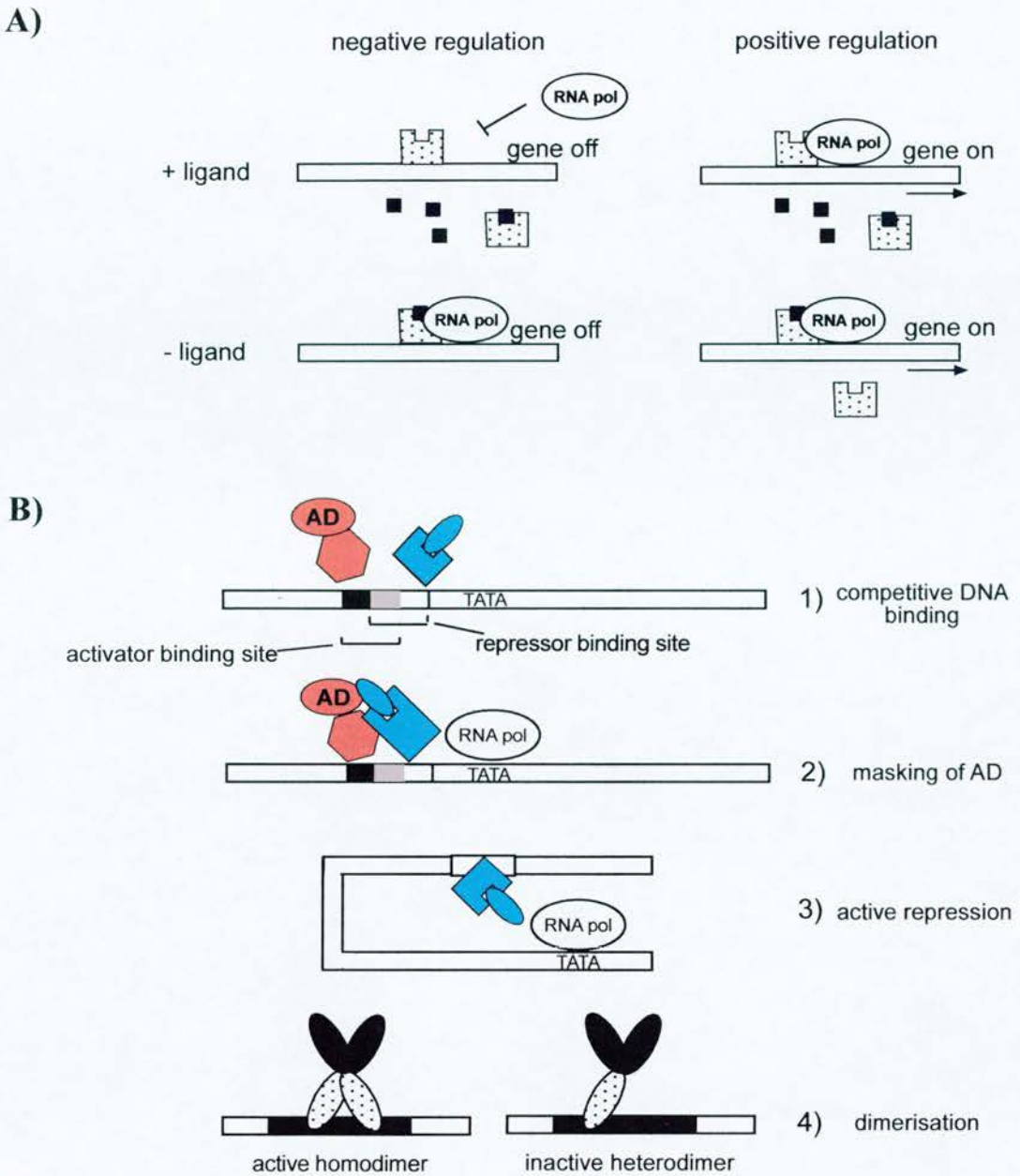


Figure 1.1 The mechanisms by which specific gene regulatory proteins control transcription

A) Positive and negative gene regulation in prokaryotes. The addition of ligand (black box) can switch off genes (negative regulation) either by removal of a gene activator protein (grey box) or by causing a gene repressor protein to bind (grey box), where it blocks the binding of RNA polymerase (steric hindrance). The addition of ligand can also switch a gene on (positive regulation) by removing a gene repressor from DNA or by causing a gene activator protein to bind, both of which enhance transcription by RNA polymerase. **B)** Negative regulation in eukaryotes. Gene activator (red) and repressor (blue) proteins may compete for binding to the same regulatory DNA sequence (1). Masking of the activation domain (AD) by a repressor protein can prevent it from contacting the transcription machinery (2). This is a form of steric hindrance. Specific protein-protein interactions can interfere with the assembly of the basal transcription machinery (3). Long-range cis-acting DNA elements may be involved in this active repression model. Heterodimerisation by truncated transcription factors (black oval) with no DBD (grey oval) can inhibit otherwise active homodimers (4).

1.1.1 The modular design of transcriptional regulators and their affect on gene regulation

Many DNA binding proteins involved in gene regulation, are composed of at least two distinct modules. One is always a DNA-binding domain (DBD) and proteins may contain one or more of these. Other conserved modules govern specific protein-protein interactions functions. In yeast, the acidic activation domain of the GAL4 DNA binding protein works by accelerating the assembly of the basal transcription machinery at the promoter (Lin and Green, 1991). Other conserved modules found in association with DBDs are responsible for mediating interactions with organic ligands or other proteins. For example, the holi domain, found in the glucocorticoid (GR) and retinoic acid (RAR) receptors is responsible for binding to the respective hormone ligand (Govindan, 1990). The SCAN domain (Williams *et al.*, 1995), also known as the leucine-rich domain (LeR) (Pengue *et al.*, 1994) serves as a protein-protein interaction interface for proteins not necessarily associated with either gene activation or repression (reviewed by Collins *et al.*, 2001).

1.1.2 DNA-binding domains

Examples of DBDs include the helix-turn-helix (HTH), zinc finger (ZF), leucine zipper (LZ) and helix-loop-helix (HLH) motifs. All families are represented in prokarya and eukarya alike, with the first structural evidence for a prokaryotic ZF protein (SmtA) being recently described (Blindauer *et al.*, 2001). In eukarya, the ZF and the HTH DBDs are the most common (Alberts *et al.*, 1994). Unlike the LZ and HLH motifs, they do not require protein dimerisation. The ability of some transcription factors to dimerise adds to the level of control and specificity they exert. The LZ is so called because of the way two monomers are held together by interactions between the hydrophobic amino acid side chains of leucine residues. LZ homodimers bind to symmetric DNA sequences via a series of basic residues adjacent to the LZ, but when two different monomers combine to form a heterodimer, a hybrid DNA sequence is recognised. The way dimerisation can inhibit gene expression is described in more detail in Figure 1.1B.

1.1.3 The ZFP transcription factor factor family

The term ZF was first used to describe the 30 amino acid long repeat sequence motif found in the TFIIIA transcription factor (Miller *et al.*, 1985). A property unique to the ZF DBD is its ability to bind RNA as well as DNA (Miller *et al.*, 1985; Caricasole *et al.*, 1996). TFIIIA is a ZF DBD protein (ZFP) that binds to 5S DNA as well as 5S ribosomal RNA but different regions of the protein are responsible for binding to each (Theunissen *et al.*, 1992). When the structure of a single ZF motif from the *Xenopus* protein Xfin was solved using NMR (nuclear magnetic resonance), it revealed each ZF motif to be folded around a central zinc ion to form an independent mini-domain (Lee *et al.*, 1989) that "grips" the DNA, as "fingers" of a hand might grip an object (reviewed by Klug and Schwabe, 1995).

1.1.3.1 C2H2-type ZFPs

The ZFP super-family can be sub-divided into similar but distinct sub-families, which have a preference to bind either nucleic acids, proteins or both. By far the most abundant sub-family are those proteins that contain the classical C2H2-type repeats similar to the repeats found in TFIIIA. The human proteome is predicted to contain 1167 proteins that harbour C2H2-type motifs ZF (InterPro proteome analysis, EMBL-EBI, January 2005), and therefore constitute one of the largest gene families in higher eukaryotes. *Drosophila* tramtrack (TTK) is an example of a C2H2-type zinc finger (Fig. 1.2A). In this repeat, one Zn atom is coordinated by the tetrahedral arrangement of two cysteine (C) and two histidine (H) residues. The Zn atom is buried in the interior and stabilises the module by binding a pair of cysteines from the β-hairpin and a pair of histidines from the α-helix (Fig.1.2A) (Klug and Schwabe, 1995). Residues from the N-terminal part of the α-helix of each finger form hydrogen bonds with three base pairs (bp) of DNA in the major groove, and so are responsible for the DNA/RNA binding specificity associated with ZFPs (Theunissen *et al.*, 1992). Multiple residues of each finger also make contact with the sugar-phosphate backbone of DNA. Most ZFP genes encode multiple tandem repeats of the ZF unit (up to 37 in Xfin, De Lucchini *et al.*, 1991) (Fig. 1.2B). Adjacent ZFs have been confirmed as having little or no interaction between them but they are combined as whole modules to expand the potential number of target

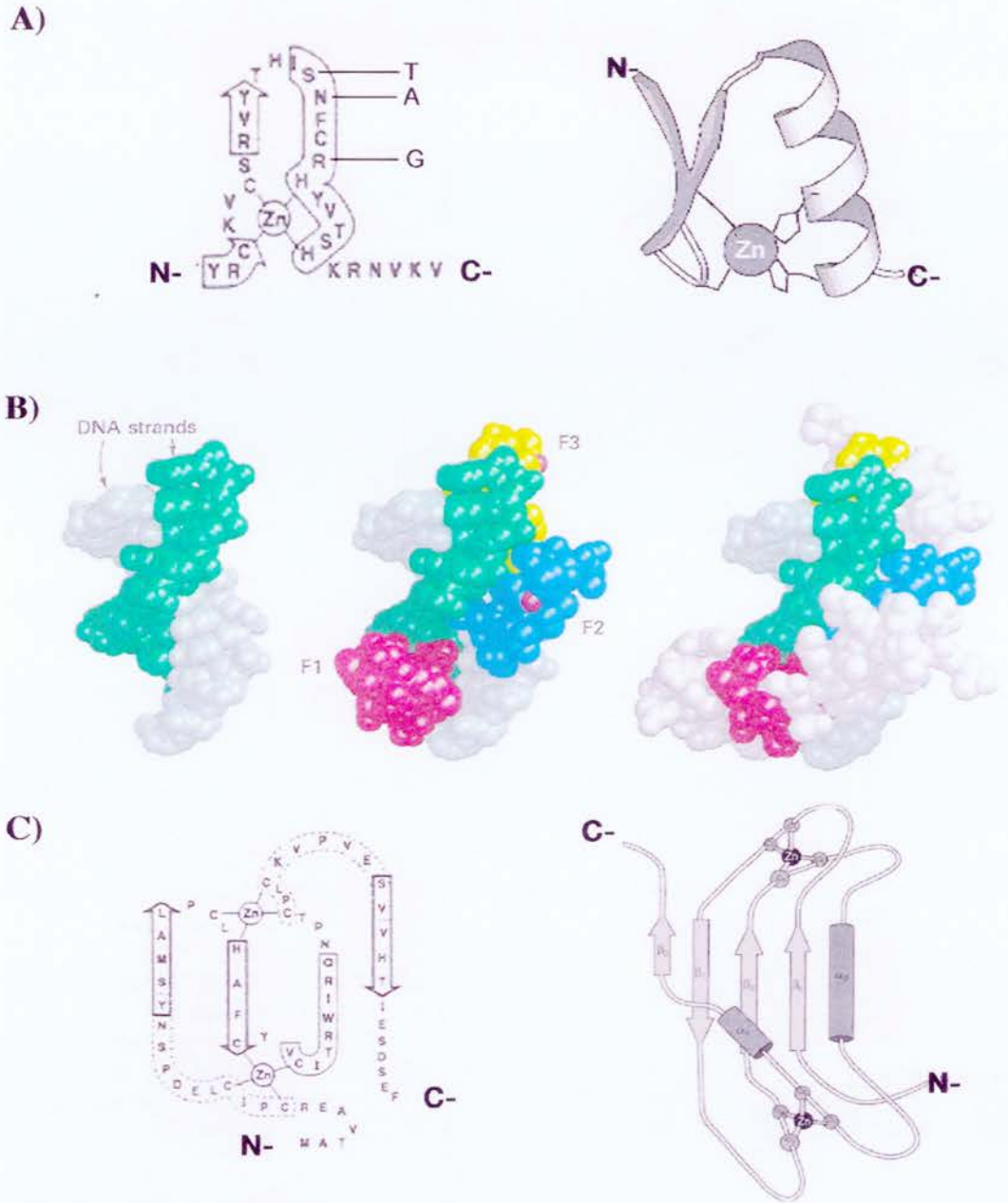


Figure 1.2 The structure of some ZF DBDs.

A) The classical C2H2-type ZF depicted on the left (adapted from Klug and Schwabe, 1995) is from the *Drosophila* protein, tramtrack (TTK) . Two cysteine residues (C) at the amino end (N-) and two histidine residues (H) at the carboxy end (C-) coordinate the Zn ion. The amino acids that contact DNA are indicated. The last C residue and the first H residue are separated by a region of 12 residues. The C2H2-type ZF can be described as a "mini-globular"domain with a hydrophobic core and with polar side chains on surface. The hydrophobic core can maintain the structure of the ZF as it binds into the major groove of DNA. The 3-D structure of a C2H2 ZF is shown on the right. Helical regions are boxed. β-strands are enclosed in arrows and the amino and carboxyl-termini are marked. **B)** The 3-D structure of Zif268 bound to DNA. *Left*, The two strands of DNA are shades of green. *Middle*, DNA with bound recognition helices (F1, F2 and F3) and Zn ions. *Right*, Entire protein-DNA complex **C)** The RING finger (C3HC4) has a single structural domain that binds two Zn ions in independent, but almost symmetrical binding pockets (adapted from Barlow *et al.*, 1994). There are examples of RINGs that have a cysteine or histidine substituted with other metal binding residues, such as aspartate or threonine (reviewed by Saurin *et al.*, 1996). Residues related by an approximate twofold symmetry are outlined with a dotted line. The 3-D structure of the PML RING finger is depicted on the right (Borden *et al.*, 1995).

genes they can bind to. A target site selection model for ZFPs would benefit the study of this large protein family, but remains a challenge for numerous reasons, one of which is the way chromatin can affects target site selection. Chromatin structure both facilitates and impedes binding of transcription factors. TBP (TATA box binding protein) is an example of a transcription factor that cannot bind to nucleosomal DNA even when the TATA box is exposed on the surface of the histone octamer (Imbalzano *et al.*, 1994). However, the winged helix transcription factor HNF3, forms a much more stable complex with nucleosomes than with naked DNA (Cirillo *et al.*, 1998). 'Designer ZFPs' have been constructed and can bind to specific target DNA sequences (Papworth *et al.*, 2003). ZF peptides are also being used as a useful tool to study the role of specific binding sites in the chromatin environment (reviewed by Segal, 2002). In particular, some 'designer ZFs' are being used as novel therapeutics and some can regulate therapeutically relevant endogenous genes.

The vast majority of C2H2-type ZFs are classified as Krüppel-type on the basis of the fact that they also share a conserved stretch of seven amino acids (the H/C link) connecting multiple, tandem repeats of the ZF domain, as found in the *Drosophila* protein, Krüppel (Rosenberg *et al.*, 1986). DNA binding assays with artificial ZFPs created from the Krüppel-family member, Sp1, shows this linker sequence also has a significant effect on DNA recognition (Nagaoka *et al.*, 2001).

1.1.3.2 RING finger family

Another sub-family of ZFPs constitute those proteins containing a RING (really interesting new gene) finger. The RING domain is the 9th most abundant motif in the human protein and is found in species ranging from viruses to plants (Stone *et al.*, 2005). There are 36 known nuclear proteins with RING fingers (Nuclear Protein database, NPD, section 2.11). The 3-D structure of the Vmw110 RING finger from equine herpes virus IE110 was the first to be structurally characterised (Barlow *et al.*, 1994). It is a member of the C3HC4 sub-class, which tetrahedrally coordinate two zinc atoms into an autonomously folded domain (Fig. 1.2C). The RING finger of PML is also a member of this sub-class, but the spacing in the RING consensus

sequence and structures around the second zinc-binding site indicate that this motif is highly variable (Borden *et al.*, 1995).

Many unrelated proteins that form supramolecular complexes have RING domains such as PML (promyelocytic leukaemia protein), BRCA1 (breast cancer associated gene 1), BARD1 (BRCA1-associated RING domain) and KAP-1 (KRAB associated protein-1), also known as TIF1β (transcription initiation factor 1β). Some RINGS can self associate as well as bind to other RINGS, but they also have roles in other protein-protein interactions, protein-DNA interactions and they also have catalytic activity in ubiquitin conjugation (Lorick *et al.*, 1999; Seeler *et al.*, 2001). The RING motif is often found in proteins in conjunction with other domains (see section 1.1.5.1). The E3-ubiquitin ligase activity of RING1A and RING1B has recently been shown to ubiquitinate H2A on lysine residue 119 (section 1.2.2.3). Since Ring1A and Ring1B are polycomb group proteins (PcG) (del Mar Lorente *et al.*, 2000; Voncken *et al.*, 2003) and ubiquitinated H2A is found on the inactive X chromosome (Xi) in female mammals (de Napoles *et al.*, 2004), this suggests they are involved in chromatin-mediated heritable gene silencing.

1.1.4 The KRAB-ZFP transcription factor family

In 1991, the conserved KRAB (Krüppel-associated box) domain was identified in human, mouse and Xenopus Krüppel-type ZFPs by sequence analysis (Bellefroid *et al.*, 1991; De Lucchini *et al.*, 1991). It was reported that about one third of Krüppel - type ZFPs contained this domain, and to date, the KRAB-ZFPs remain the largest class of transcriptional regulators in the mammalian genome (reviewed by Urrutia, 2003). So far, they are restricted to the genomes of tetrapods (land-living vertebrates), as members of this family have not been identified in yeast, *Arabidopsis*, *Drosophila* or other vertebrates such as fish (Mark *et al.*, 1999; Urrutia, 2003). Considering there are 355 KRAB box containing C2H2-type ZFPs in the human genome (InterPro, proteome analysis, January 2005) and that the initial characterisation of such proteins was carried out over a decade ago (Bellefroid *et al.*, 1991), comparatively little is still known of their specific biological functions. The small number of biological roles already identified for KRAB-ZFPs is a result of

their chance involvement in other biological pathways (Casademunt *et al.*, 1999; Zheng *et al.*, 2000; Skapek *et al.*, 2000; Li *et al.*, 2003; Hennemann *et al.*, 2003; Nielsen *et al.*, 2004). The deficit of KRAB-ZFP target DNA sequences contributes to this lack of physiological knowledge.

The KRAB domain is a highly conserved 75 amino acid motif found almost exclusively in the N terminus of C2H2 ZFPs (Bellefroid *et al.*, 1991). The KRAB-ZFPs can be sub-divided into four families based on the composition of the KRAB box (Fig. 1.3A). All sub-families contain a KRAB A box which is encoded by a single exon (Bellefroid *et al.*, 1991). The remaining families contain a KRAB A box in conjunction with either a classical B, highly divergent b, or the newly identified C box (Mark *et al.*, 1999; Looman *et al.*, 2004). Each of these boxes is also encoded by individual exons and they are predicted to fold into charged amphipathic helices (Bellefroid *et al.*, 1991; Looman *et al.*, 2004). KRAB A family members often result by alternative splicing out of the B exon (Bellefroid *et al.*, 1993). Additional families include those KRAB-ZFPs known to contain more than one KRAB box (Looman *et al.*, 2004). Still others are often found to contain a SCAN domain (Pengue *et al.*, 1994; Casademunt *et al.*, 1999). The SCAN domain consists of at least 87 residues known to mediate homo- and hetero-oligomerisation possibly with other SCAN domain proteins (Collins *et al.*, 2001; Honer *et al.*, 2001). The ZF repeats in all KRAB-ZFP families are located together at the C-terminus and are encoded by one exon (Bellefroid *et al.*, 1993; Villa *et al.*, 1996), including ZNF91 and Xfin, which contain 35 and 37 contiguous repeats respectively.

1.1.5 KRAB-ZFP interaction with the KAP-1 co-repressor

The KRAB domain was shown to function as a potent DNA binding dependent transcriptional repressor, when fused to a heterologous DBD from the yeast GAL4 protein (Margolin *et al.*, 1994; Witzgall *et al.*, 1994). In this experiment the KRAB A box was deemed necessary and sufficient for repression. The KRAB B box was subsequently found to contribute to transcriptional repression but only to a lesser extent (Margolin *et al.*, 1994; Witzgall *et al.*, 1994; Pengue et al., 1994), whilst the

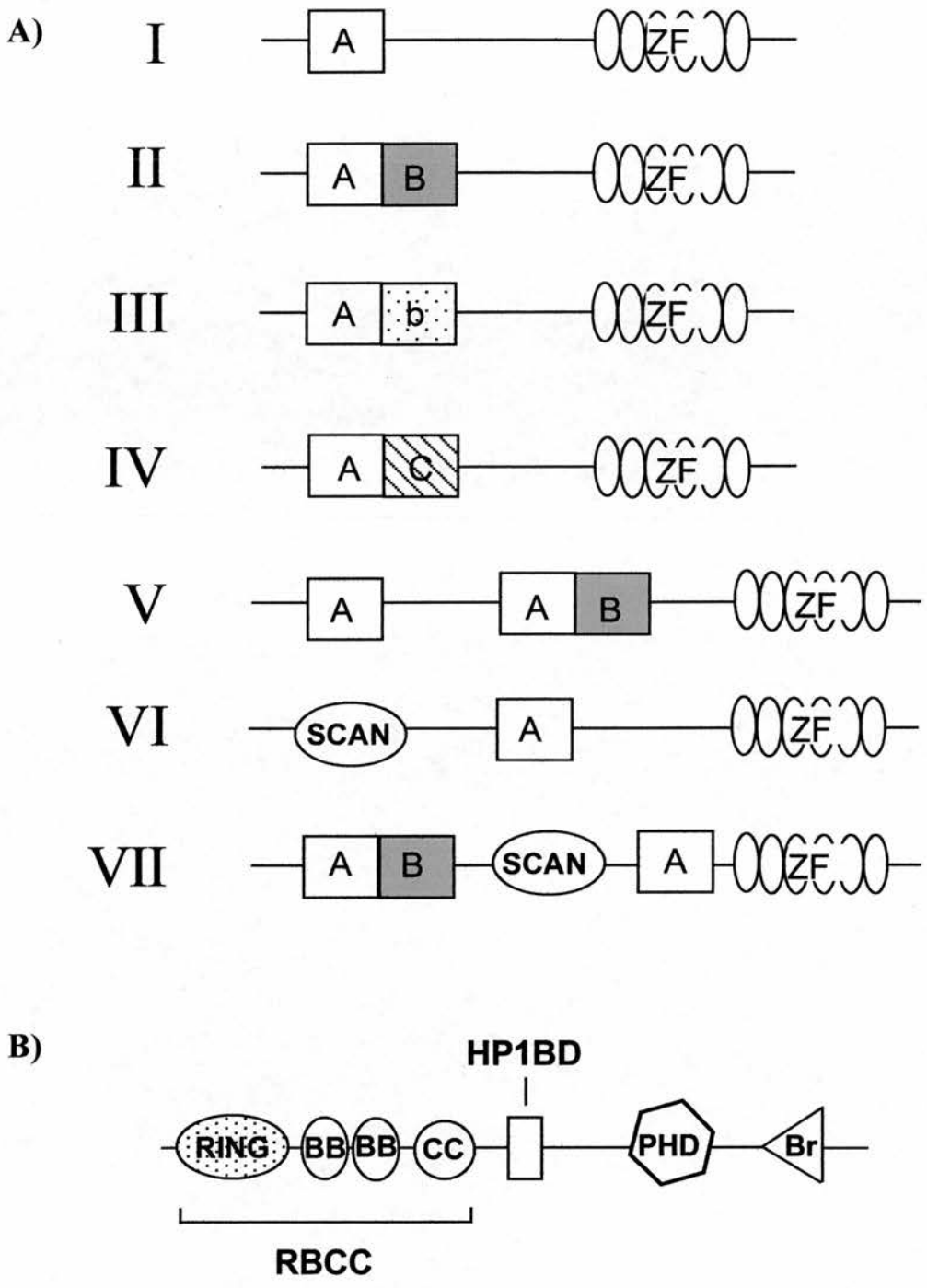


Figure 1.3 KRAB-ZFP sub-families

A schematic representation of the structures of the KRAB-ZFP sub-families found to date (**A**) and KAP-1 (**B**). *A, B and C* are the KRAB A, B and C domains respectively. The ZF (zinc finger) motif in each sub-family is represented as a group of six. The RING, BB (B boxes) and CC (coiled coil) regions are collectively known as the RBCC motif. *HP1BD*, HP1 binding domain; *PHD*, plant homeodomain; *Br*, bromodomain. Diagram not to scale.

highly divergent b box was found to have no affect on transcriptional repression (Vissing *et al.*, 1995; Åbrink *et al.*, 2001). These experiments also showed that the recently identified KRAB C box, like KRAB b, does not contribute to repression, but it does appear to strengthen the interaction between the KRAB domain and KAP-1 (see below) (Looman *et al.*, 2004).

To identify other proteins responsible for transcriptional repression by KRAB-ZFPs, various routes were taken. Evidence for a titratable cellular factor required for KRAB-domain mediated repression was obtained by squelching (inhibition of transcriptional repression) of GAL4-KRAB-mediated repression with a non-DNA-bound KRAB domain (Friedman *et al.*, 1996). The cellular factor found to be responsible, KAP-1 (TIF1β), was then identified by affinity chromatography. The same result was obtained with yeast two-hybrid experiments using the KRAB domain (Moosmann *et al.*, 1996).

1.1.5.1 The TIF family of proteins

Four mammalian TIFs have been characterised including TIF1α, KAP-1, TIF1-γ and TIF1-δ (Le Douarin *et al.*, 1995; Ryan *et al.*, 1999; Venturini *et al.*, 1999; Khetchoumian *et al.*, 2004). All repress genes when tethered to a promoter and all contain an RBCC motif at the N-terminus and a PHD (plant homeodomain) and bromodomain at the C-terminus (Fig. 1.3B). The RBCC/TRIM (tripartite) motif is composed of a C3HC4 RING finger (R), two additional Zn binding Cys/His-rich clusters (the B-boxes) (Borden *et al.*, 1993), and a putative α-helical coiled coil region (CC). Other proteins containing this motif, or parts of, include Mid1, which exhibits E3-ligase activity and is the cause of Opitz syndrome (Trockenbacher *et al.*, 2001), HLS5, which is a candidate tumour suppressor (Lalonde *et al.*, 2004) and MuRF-1, which ubiquitinates troponin I (Kedar *et al.*, 2004).

In addition to the domains mentioned above, TIF1α, -β and -δ also contain a central penta-peptide sequence (PXVXL, where X is any amino acid), responsible for HP1 binding to the HP1 binding domain (HP1BD). TIF1α has been shown to associate with a wide variety of ligand-bound nuclear hormone receptors and to function as a

co-activator for retinoic acid receptors (Le Douarin *et al.*, 1995). The nuclear receptor interaction domain (NRBD) is adjacent to the HP1BD. Unlike TIF1α and -γ, KAP-1 does not have a NRBD and does not interact with nuclear receptors, but TIF1α and KAP-1 can both bind to the KRAB box. TIF1γ and TIF1δ neither bind to nuclear receptors, nor to KRAB-ZFPs (Venturini *et al.*, 1999; Åbrink *et al.*, 2001; Khetchoumian *et al.*, 2004). Instead it is thought that these TIFs function via oligomerisation and/or by interaction with other proteins. All homologues can homodimerise but TIF1α and -γ have also been shown to heterodimerise (Peng *et al.*, 2002) and this heterodimerisation is functionally significant in childhood papillary thyroid carcinomas (Klugbauer and Rabes, 1999). TIF1δ is unique among the TIF1 family in that its expression is largely restricted to the testes where it associates preferentially with HP1γ (Khetchoumian *et al.*, 2004). Gene targeting experiments have only been reported for KAP-1 which was found to be embryonic lethal and required for post-implantation embryogenesis and mesoderm induction (Cammas *et al.*, 2000). TIF1γ has recently been reported as an essential regulator of embryonic and adult haematopoiesis in Zebrafish and whilst a TIF1α orthologue has also been identified, a KAP-1 orthologue has not (Ransom *et al.*, 2004). Therefore, KAP-1 and the KRAB domain still remain restricted to tetrapods. A single TIF1 homologue, (Bonus) has been found in the *Drosophila* genome (Beckstead *et al.*, 2001; Beckstead *et al.*, 2004) and shares with TIF1α the ability to interact and negatively regulate with nuclear receptor dependent transcription.

KAP-1 was also cloned as an interactor, along with TIF1α, of the HP1 (heterochromatin protein 1) proteins (see section 1.3.1) demonstrating a potential role in repressive chromatin structures (Le Douarin *et al.*, 1996b). Similar to the other TIFs, KAP-1 contains a Cys/His cluster termed the PHD domain (plant homeodomain) and a bromodomain at its C-terminal. These domains are found in many proteins implicated in chromatin-mediated transcriptional regulation (Aasland *et al.*, 1995; Syntichaki *et al.*, 2000). KAP-1 interaction with chromatin remodelling proteins is discussed in section 1.3 and the effect of the KRAB-KAP-1 interaction on chromatin structure is discussed in more detail in section 1.3.4.

1.1.5.2 Transcriptional repression by the KRAB-KAP-1 complex

KRAB-mediated silencing requires binding to the KAP-1 co-repressor (Agata *et al.*, 1999; Peng *et al.*, 2000) and when tethered to DNA, KAP-1 and KRAB-ZFPs can repress transcription from both proximal and remote promoter positions (Pengue *et al.*, 1994; Moosmann *et al.*, 1996) (Fig. 1.1B, 3). Despite KAP-1 having the ability to co-ordinate chromatin structure, KRAB-KAP-1 mediated regulation is also reported to be by active repression (Fig. 1.1B, 3). Evidence for this is that KRAB-KAP-1 mediated repression is efficient even in short-term transient transfection assays before the bulk of reporter DNA templates would be fully assembled into chromatin (Pengue *et al.*, 1994; Agata *et al.*, 1999). Active repression may be achieved by inhibitory protein-protein interactions but evidence that the bromodomain of TIF1α (and KAP-1, Ryan *et al.*, 1999) possesses intrinsic protein kinase activity capable of autophoshorylation and phosphorylation of transcription factors suggests that direct interaction may not be necessary (Fraser *et al.*, 1998; Ryan *et al.*, 1999). The KRAB domain is able to repress transcription by Pol II and Pol III, but not by Pol I, suggesting that KRAB exerts some of its repression activity by interfering with components of the basal transcription machinery. However, it is possible that Pol I could simply ignore any chromatin modifications.

Oligomerisation of KRAB-ZFPs has been shown to play a role in transcriptional regulation (Li *et al.*, 2003; Tan *et al.*, 2004b). A KRAB and SCAN-only containing protein (minus ZF domain), VHLaK, was recently shown to function as a transcription repression module via homo-oligomerisation of the SCAN domain (Fig. 1.1B, 4). KRAB-ZFPs are also likely to compete with other transcriptional regulators for target sequences and so gene expression is also effected in this way (Jheon *et al.*, 2001; Tanaka *et al.*, 2002; Jheon *et al.*, 2003) (Fig. 1.B, 1).

Evidence that the KRAB-ZFP class of transcriptional repressors appears to mediate repression via all of the mechanisms described in Figure 1.1B has been presented. However, DNA in cells is assembled into chromatin, which imposes a very poorly understood set of rules on the binding and action of transcriptional regulators. Thus,

the fifth mechanism of repression utilised by transcription factors is the interaction with co-repressor and co-activator proteins that mediate epigenetic re-organisation of chromatin into an active/inactive form. The KRAB-KAP-1 interaction is a good illustration of how a transcription factor and co-repressor initiates the specific recruitment of chromatin-associated enzyme complexes for the subsequent modification of chromatin structure. It is now clear that KAP-1 acts as a scaffold to bring together various proteins that modify chromatin structure and so mediate transcriptional repression (section 1.3.4). In order to understand this level of regulation we must first understand the molecular basis of chromatin structure.

1.2 Chromatin structure

Chromatin is composite of DNA, histones and non-histone proteins. Histones are the most highly conserved proteins in the eukaryotic genome (Wolffe, 1995). They are small (11-16 kDa) positively charged (lysine and arginine rich) proteins. 146 bp of negatively charged DNA is wrapped around a core octameric complex formed by dimers of histone proteins H2A, H2B, H3 and H4 to form a nucleosome (Luger *et al.*, 1997). The C-terminal histone fold domains of core histones are highly structured and found on the inside of the nucleosome, but the N-termini are flexible, and protrude from the nucleosome where they can make contacts with DNA, nucleosomes, and other non-histone proteins. In this way, nucleosomes and chromatin fibres are not structurally inert. Linker histones bind to DNA between nucleosomes with a stoichiometry of one H1 protein per nucleosome and interact with the core H2A sub-units. DNA folds into a 30 nm chromatin fibre (Widom, 1998; Zlatanova *et al.*, 1999) and other higher order structures based on modification of the nucleosomes and of the DNA itself. It is now known that chromatin can fold into a 30 nm fibre in the absence of H1 and it is the core histone tails that are critical for chromatin condensation (Carruthers *et al.*, 1998). The compaction and decompaction of chromatin fibres makes regions of chromatin refractory or acquiescent to processes such as transcription, replication, recombination and repair.

1.2.1 DNA methylation

In animals, DNA is methylated on cytosine residues of CpG dinucleotides at the C5 position. This modification is essential for mammalian development and in control of gene expression (Okano *et al.*, 1999). In the adult vertebrate genome, 60-90% of all CpG dinucleotides are methylated. The remaining non-methyated CpGs (~15% in humans) are found in CpG islands, which usually include functional promoters and are predominantly found in early replicating (R-bands) of the genome (Craig and Bickmore, 1994).

DNA methylation is normally associated with gene silencing (Bird and Wolffe, 1999), X-inactivation and genomic imprinting (Ng and Bird, 1999). Three DNA methyltransferases (DNMT1, -3a, and -3b) have been identified (Bestor *et al.*, 1988; Okano *et al.*, 1998). This DNA mark plays a fundamental role in epigenetic gene silencing as the methylation mark can be copied after replication, resulting in heritable changes in chromatin structures. Imprinted genes are associated with differentially methylated regions (DMRs) in which maternal and paternal copies of the gene are methylated differently (Neumann *et al.*, 1995) and mono-allelic expression results as a consequence.

DNA methylation might influence gene expression in two ways. It might directly repress transcription by blocking the binding of transcription factors. Secondly, it can create a binding site for repressor proteins. A family of methyl-CpG binding proteins (MBDs) have been identified on the basis of a conserved methyl-binding domain (MBD) (Hendrich and Bird, 1998). The founder member, MeCP2, is a transcriptional repressor that recruits repression complexes containing histone deacetylases (section 1.2.2.1, Nan *et al.*, 1998). MBD1 to –3 have also been shown to be potent transcriptional repressors. MBD2 is a component of the MeCP1 repressor complex and interacts directly with Sin3A (Boeke *et al.*, 2000). MBD4 is a DNA glycosylase which repairs G:T mismatches at CpG sites (Hendrich *et al.*, 1999). MBD3 has been shown to be a member of the NuRD complex (Zhang *et al.*, 1999). More recently, MBD1 has been shown to recruit the histone methyltransferase (HMT) SETDB1 to the large subunit of CAF-1 (chromatin

assembly factor 1) to form an S-phase-specific complex that facilitates methylation of H3-K9 during replication-coupled assembly (Sarraf and Stancheva, 2004). SETDB1 also has a CpG DNA methyl-binding domain, which if functional, could maintain H3-K9 methylation at a target locus (Schultz *et al.*, 2002; Ayyanathan *et al.*, 2003). Dnmt1, which interacts with PCNA (an auxiliary component of the replication machinery), is involved in the heritability of DNA methylation patterns and this activity could supplement H3-K9 methylation by SETDB1.

Other experiments have suggested a close link between DNA methylation and histone modifications. In *Neurospora*, DNA methylation is controlled by dim-5, the H3-K9 histone methyltransferase (HMT) protein (Tamaru and Selker, 2001). HMT activity of the KRYPTONITE protein is also involved in maintenance of DNA methylation in *Arabidopsis* (Jackson *et al.*, 2002). MBD1 (methyl-CpG binding domain 1) has been found to interact with the Suv39h1-HP1 heterochromatic complex (section 1.3.1) for DNA methylation-based transcriptional repression. Furthermore, MBD1 links to histone deacetylases through Suv39h1 and suggests the presence of a pathway from DNA methylation to the modification of histones for epigenetic gene-regulation (Fujita *et al.*, 2003). This data presents the growing relationships between DNA modifications, chromatin modifications and gene expression.

1.2.2 Core histone modifications

Modifications of the core histones are associated with transcriptionally 'active' or 'silent' chromatin. The N-terminal tails are subject to post-translational modifications including acetylation, methylation and phosphorylation whilst ubiquitination occurs mainly at the C-terminus (Fig. 1.4). Such modifications rework the chromatin fibre by altering the electrostatic charge of the histones and/or through creating binding sites for non-histone proteins.

A "histone code hypothesis" was proposed by Strahl and Allis (2000) which suggested that combinations or sequences of histone modifications results in the signalling of specific downstream events by recruitment of transcriptional activators

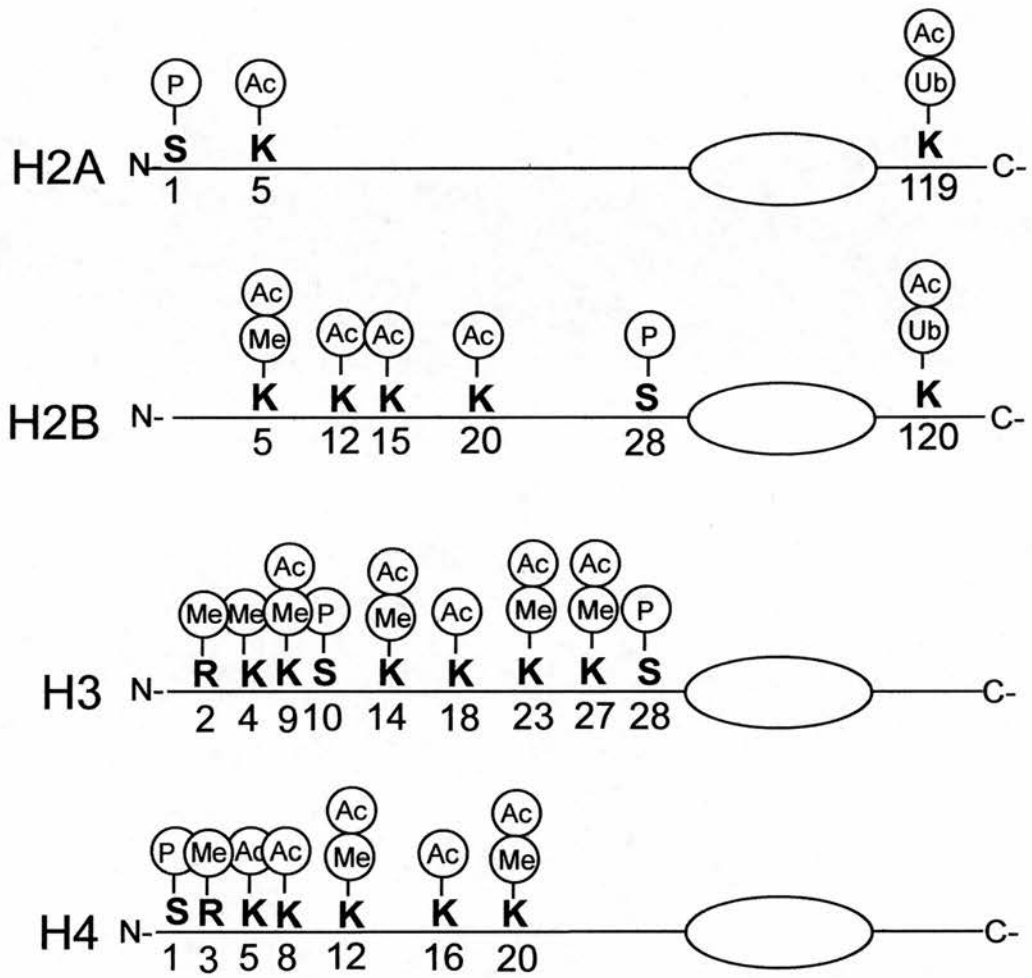


Figure 1.4 Histone modifications

Core histones consist of a central globular domain (oval) and N- and C- terminal tails that protrude outwards from the globular core. Specific residues at the N- and C-terminal tails can be chemically modified by methylation (Me), acetylation (Ac), phosphorylation (P) and ubiquitination (Ub) and a subset of these marks are illustrated. *H*, histone protein; *K*, lysine; *S*, serine; *R*, arginine.

or repressors. This original hypothesis has been extended to include groups of particular histone modifications that work in cohort to provide chromatin signals that either permit or silence gene expression (Fischle *et al.*, 2003b). It is proposed that histone modifications primarily affect the interaction with other proteins that concurrently influence chromatin structure and hence DNA accessibility, rather than histone modifications altering the nucleosome or chromatin structure directly.

1.2.2.1 Histone acetylation

In 1964, Alfrey and colleagues proposed a link between acetylation of lysine residues within the N-terminal domains of core histones (particularly H3 and H4) and the regulation of transcription (see review by Peterson *et al.*, 2002). It is unclear as to how histone acetylation specifically alters or influences the biophysical properties of chromatin. It could be that acetylation of N-terminal tails results in a weakened interaction with the DNA backbone which would facilitate the interaction of transcription factors with their target site (McGhee *et al.*, 1983; Wolffe and Hayes, 1999) but there is evidence to suggest that this is not the case (Marmorstein, 2001). Instead acetylation may be required for maintenance of an active chromatin state rather than its establishment (Cavalli and Paro, 1999), which may be achieved by acetylated histones creating a binding site for bromodomain-containing transcriptional regulators (Fischle *et al.*, 2003a). It should also be noted that many transcriptional regulators, for example, p300/CBP and components of the basal transcription machinery are also acetylated, thus acetylation has diverse affects on gene regulation (Li *et al.*, 1998; Li *et al.*, 1999).

The discovery that co-activator complexes required for transcriptional activation also contain histone acetyltransferase (HAT) activity was first observed in yeast (Brownell *et al.*, 1996; Grant *et al.*, 1997). There is a high degree of sequence similarity between members of each of the HAT families (reviewed by Kuo and Allis, 1998) and they often encode two conserved domains, the acetyltransferase domain and the bromodomain (Dhalluin *et al.*, 1999; Marmorstein *et al.*, 2001). The bromodomain of the yeast Gcn5 HAT binds to acetylated lysines and is not required for *in vivo* Gcn5-mediated histone acetylation but is required for subsequent Swi2-

dependent nucleosomal remodelling (section 1.3.3) and subsequent transcriptional activation, indicating that it may constitute a targeting step for events following histone acetylation (Syntichaki *et al.*, 2000). Many human HATs have been described and they are components of multi-subunit complexes. These include PCAF (the Gcn5 human homologue), p300/CBP and TAF250 that act locally at gene promoters and modulate gene specific transcription (Owen *et al.*, 2000; Jacobsen *et al.*, 2000).

There are three main classes of HDACs identified. Class I and class II are based on their similarity to the *S. cerevisiae* enzymes Rpd3p or Hda1p respectively, whilst class III include the NAD+-dependent Sir2 family members (reviewed by Kraus and Wong, 2002). In humans, HDAC 1, 2 and 3 are members of class I whilst HDACs 4, 5 and 6 are class II members. HDACs 1 and 2 are found in the major chromatin remodelling complexes, NuRD (for nucleosomal remodelling and histone deacetylation, also known as Mi-2) and SIN3 (Bird and Wolffe, 1999; Knoepfler and Eisenman, 1999). In addition to HDAC activity, NuRD contains ATP-dependent nucleosomal remodelling activity, MBD3, and the histone binding proteins RbAp46 and RbAp48, the latter two of which are also found in the SIN3 complex (reviewed by Ahringer, 2000). SIN3 has been extensively studied with respect to its involvement in repression by nuclear hormone receptors. It physically links to the co-repressors N-CoR and SMRT (Heinzel *et al.*, 1997; Nagy *et al.*, 1997) suggesting that the repression by these co-repressors is linked to deacetylation of histones. Consistent with this, transcriptional repression by nuclear hormone receptors is blocked by TSA (Heinzel *et al.*, 1997).

The distinction between histone hypoacetylation and hyperacetylation, is determined by the number of acetylated lysine residues per H4 molecule (Zhang *et al.*, 2002). Three or more acetylated lysines is indicative of hyperacetylation and in general hyperacetylation correlates with transcriptional activity while hypoacetylation correlates with transcriptional repression (reviewed by Marmorstein, 2001). The acetylation pattern for specific histone tails and their relation to gene activity has been described for the human interferon-β locus (Agalioti *et al.*, 2002), mouse β-

globin (Bulger *et al.*, 2003) and mouse insulin (Chakrabarti *et al.*, 2003). The inactive X is hypoacetylated (Plath *et al.*, 2002).

Although it has been suggested that hyperacetylation of histones by inhibition of histone deacetylases (HDACs) with TSA (Trichostatin A) causes gross remodelling of nuclear structure (Taddei *et al.*, 2001), recent evidence suggests otherwise (Gilchrist *et al.*, 2004).

KAP-1, may also mediate its repressive effects through deacetylation. Evidence for this stems from the fact that TSA partially inhibits the silencing activity of KAP-1 (Nielsen *et al.*, 1999) and KRAZ1 (Matsuda *et al.*, 2001) when fused to the GAL4 DBD. However, although a small amount of KAP-1 is known to be complexed to Mi-2α (section 1.3.3) (Schultz *et al.*, 2001), TSA could still partially inhibit deacetylation in C-terminal mutants that are devoid of the Mi-2α-KAP-1 interaction (Matsuda *et al.*, 2001). In another *in vitro* study with stable or transiently transfected reporter templates, HDAC inhibition did not alter KRAB-KAP-1 mediated repression nor did they show KAP-1 to be in complex with SIN3 (Lorenz *et al.*, 2001), however the repression activity of TIF1α was severely affected.

1.2.2.2 Histone methylation

Core histones can also be methylated at lysine and arginine residues (Fig. 1.4) and the interplay between lysine methylation and acetylation influences transcription. Each lysine residue can contain up to 3 methyl groups. Methylation does not alter the overall charge of histone tails, but the addition of methyl groups does increase their affinity for DNA, and also increases the resistance of DNA to trypsin digestion (reviewed by Rice and Allis, 2001). Over the past four years there has been an "explosion" of histone methyltransferases discovered.

Genes encoding mammalian homologues of *Drosophila Su(var)3-9* (*Suv39h1* and *Suv39h2*) were the first to be shown to encode proteins with lysine HMT activity (Rea *et al.*, 2000). Most HMTs contain the 130-amino acid SET domain (for *Su(var)3-9*, *Enhancer of Zeste*, *Trithorax*). Suv39h1 and –h2 display site-selectivity

towards H3-K9 methylation at unmodified H3 amino termini (Rea *et al.*, 2000). SETDB1 is also H3-K9 specific and facilitates the conversion of di-methylation to tri-methylation (Schultz *et al.*, 2002; Wang *et al.*, 2003). Other HMTs methylate H3 and H4 lysine residues specifically at other positions, for example, TRX (Beisel *et al.*, 2002) and ASH1 (Cao *et al.*, 2002), whilst G9a is a dual HMT that methylates lysines specifically at positions 9 and 27 of H3 (Tachibana *et al.*, 2001).

Both lysine and arginine histone methylation is involved in transcriptional regulation. CARM1 (co-activator-associated arginine methyltransferase) specifically methylates arginines 17 and 26 of histone H3 but also methylates non-histone proteins such as p300/CBP. The intrinsic methyltransferase activity of CARM1 and interaction with SRC protein stimulates nuclear hormone receptor activity in mammalian cell-based reporter gene assays (Chen *et al.*, 1999). SUV39H1, -H2 and G9a mediate transcriptional silencing, since the H3-K9 mark serves as a binding site for heterochromatin protein 1 (HP1) (section 1.3.2.1, Jacobs *et al.*, 2001). The HMT protein, Dot1 (Disrupter of telomeric silencing-1) (van Leeuwen *et al.*, 2002), is involved in silencing at telomeres, the mating type locus and rDNA in budding yeast by catalysing H3-K79 methylation. However, this protein does not contain a SET domain. *Drosophila* ESC-E(Z) catalyses H3-K27 methylation and is associated with transcriptional silencing of the *HOX* locus (Muller *et al.*, 2002) and the human homologue, EED-EZH2, is also associated with transcriptional silencing and X-inactivation in human cells (Cao *et al.*, 2002). The influence of histone methylation on transcription may not only depend on the residue methylated but also on the number of times it is methylated. For example, tri-methylated H3-K9 is enriched at pericentromeric heterochromatin whilst mono and di-methylated forms localise specifically to silent domains within euchromatin (Rice *et al.*, 2003). Similarly, tri-methylation of H3-K4 is specific for active transcription whilst di-methylation is present in both active and repressed genes (Santos-Rosa *et al.*, 2002). H3-K4 and H3-K9 methylation have been proposed to be mutually exclusive (Heard *et al.*, 2001; Boggs *et al.*, 2002). However, H3-K4 methylation has little affect on the ability of SETDB1 to methylate H3-K9, in contrast to SUV39H1 whose enzymatic activity for H3-K9 is significantly reduced by a methylated H3-K4 substrate (Schultz *et al.*,

2002). This may help to prevent the spread of transcriptionally active or silent chromatin. The co-ordination of histone methylation and acetylation on H3-K4, 9, 27 and H4-K20 suggests interplay between these epigenetic marks. H3-K9 acetylation and methylation are mutually exclusive suggesting histone deacetylation precedes methylation at these residues (Nielsen *et al.*, 2001b). Histone H4 hyperacetylation precludes H4 lysine 20 trimethylation but not mono or di-methylation (Sarg *et al.*, 2004). Recently, proteins capable of reversing the methylation mark were discovered. For example, the transcriptional co-repressor, LSD1, functions as a histone demethylase and specifically demethylates histone H3 lysine 4, which is linked to active transcription (Shi *et al.*, 2004).

1.2.2.3 Other histone modifications

Histones can also be phosphorylated on serine and threonine residues of core histones. Phosphorylation of histone H3 on serine 10 inhibits H3-K9 methylation so that HP1 can no longer bind (Rea *et al.*, 2000), but it is coupled with H3-K9 and -K14 acetylation and transcriptional activation during mitogenic and hormonal stimulation in mammalian cells (Cheung *et al.*, 2000). Aberrant H3-K9 methylation can antagonize serine 10 phosphorylation, leading to mitotic chromosome dysfunction (Rea *et al.*, 2000).

Other post-translational modifications of histones include ADP-ribosylation, glycosylation (MBD4, section 1.2.1) and ubiquitination (section 1.1.3.2). Recently, the rapid decondensation of chromatin was found to be associated with polyADP-ribosylation and important for long-term memory (Cohen-Armon *et al.*, 2004). Histone H2B was found to be ubiquitinated in *S. cerevisiae* (Robzyk *et al.*, 2000), and although polyubiquitylation has a role in protein turnover, histones are generally monoubiquitylated, a form of the modification that is not associated with protein degradation (reviewed by Osley, 2004). Ubiquitination of H2B plays an important role in the trans-histone methylation of histone H3, a modification with close ties to the regulation of gene expression.

Histone modifications provide an interface between DNA and non-histone chromatin proteins that can then affect gene transcription but there are several core histone variants that confer specialised properties to nucleosomes and are differentially expressed during development (Romano, 1992; Wolffe, 1995). A number of H2A variants are non-randomly distributed on the active and inactive X chromosome (Chadwick *et al.*, 2001). H2A.Z has been associated with transcriptional control as deletion mutants in yeast exhibit slower growth patterns (Santisteban *et al.*, 2000). The histone variant H3.3 has a conserved N-terminal tail and can replace histone H3 at sites of active transcription adding a new layer of complexity and possibilities to the regulation of transcription through changes in chromatin structure (McKittrick *et al.*, 2004).

1.3 Euchromatin and heterochromatin proteins

Cytological studies distinguished two types of chromatin based on their microscopic structural characteristics. DNA dyes, such as DAPI, highlight portions of the interphase genome that stain more intensely than the surrounding euchromatin. This constitutive heterochromatin is generally associated with regions of tandem repeats, such as at centromeres (reviewed by Richards and Elgin, 2002). Facultative heterochromatin is euchromatin that will adopt specific heterochromatic properties in a spatially and temporally controlled manner. In nucleated erythrocytes, mouse and chicken have different mechanisms of facultative heterochromatin formation. Mouse embryonic erythrocytes require HP1 proteins unlike chicken, *Xenopus* and zebrafish erythrocytes (Gilbert *et al.*, 2003). In chicken, differentiation of erythrocytes coincides with the appearance of the variant linker histone H5. The female inactive X-chromosome is another example of facultative heterochromatin formation.

Histone modifications associated with constitutive heterochromatin are hypoacetylation of histones, H3-K9 methylation, DNA methylation (reviewed by Richards and Elgin, 2002), mono-methylation of H3-K27 (Peters *et al.*, 2003) and H3-K20 tri-methylation (Schotta *et al.*, 2004). Facultative heterochromatin is not characterised by repetitive sequences but can be defined as euchromatic regions that become packaged into a heterochromatic-like form with many of the same molecular

signatures as constitutive heterochromatin at the nucleosomal level. Histone hypoacetylation and H3-K9 methylation occur on the inactive X chromosome (Peters *et al.*, 2002) but it is not depleted in acetylated H4 within the XY body during spermatogenesis (Armstrong *et al.*, 1997).

The biochemical and structural differences between eu- and heterochromatin directly influence transcriptional regulation. When the expression of a gene variegates with respect to heterochromatin, the phenomenon is called position effect variegation (PEV) (Csink and Henikoff, 1996). A euchromatic transgene can become integrated adjacent to a block of heterochromatin, and transcription may be silenced in only a subset of cells, and this state is stably inherited by their progeny leading to variegated or mosaic patterns of gene expression (Wakimoto, 1998). Many proteins that function at heterochromatin and help to establish repressive chromatin structures were identified as modifiers of PEV *Su(var)s* in *Drosophila* genetic screens. Homologues of these were then identified in mammals. Examples include, Su(var)2-5 (HP1) and the Su(var)3-9 HMTs. A new group of PEV mutants (Su(var)3-1) have recently been identified that reflect a gain-of-function and antagonize the expansion of heterochromatin (Ebert *et al.*, 2004).

The dependency of epigenetic modification of heterochromatin structure by proteins is reflected by the fact that mutations in the human *DNMT3b* gene cause ICF syndrome (Immunodeficiency-centromeric instability-facial anomalies syndrome) and patient cells show destabilisation of centromeric heterochromatin (Xu *et al.*, 1999). Similarly, loss of *DNMT1* in human cancer cells affects nuclear organisation of pericentromeric heterochromatin by decreasing H3 methylation and increasing H3 acetylation (Espada *et al.*, 2004). However, the erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin without substantially changing the structural properties of heterochromatin and there was no detectable activation of transcription (Tariq *et al.*, 2003).

1.3.1 HP1 and the propagation of heterochromatin

Mammals have three homologues of *Drosophila Su(var)*2-5, termed HP1α, -β and –δ (Singh *et al.*, 2002; Saunders *et al.*, 1993; Ye and Worman, 1996). Each isoform contains a chromodomain, which binds to H3-K9 and is essential for heterochromatin-mediated gene silencing (Bannister *et al.*, 2001; Lachner *et al.*, 2001), and a chromoshadow domain, which is involved in homo and heterodimerisation and a number of different protein interactions with transcriptional regulators such as KAP-1 and TIF1α (Le Douarin *et al.*, 1996a), SUV39H1 (Aagaard *et al.*, 1999) and lamin B receptor (Ye *et al.*, 1997). KAP-1 and SUV39H1 are often situated at heterochromatin (Ryan *et al.*, 1999; Agaard *et al.*, 1999), however TIF1α and the lamin B receptor are associated with the nucleoplasm and the nuclear periphery respectively (Remboutsika *et al.*, 1999; Stuurman *et al.*, 1998).

The dose-dependent effects of HP1 on heterochromatin-mediated gene silencing in both flies and mammals has been well documented (Eissenberg *et al.*, 1992; Festenstein *et al.*, 1999; Festenstein *et al.*, 2003). PEV is stochastic in nature due to the role HP1 plays in the variable spreading of heterochromatin and thus transcriptional silencing. The link between H3-K9 tri-methylation and HP1 directly links stable epigenetic gene silencing with heterochromatin structure. A recent model for the proposed propagation of heterochromatin initiates with methylation of H3-K27 and H3-K9 by currently unknown H3-K27 and H3-K9 mono-methylases (Fig. 1.5). Methylation of mono-methylated H3-K9 by Suv39h would then ensue with the subsequent binding of HP1α and -β to tri-methylated H3-K9. The HP1 molecules are then predicted to recruit the novel Suv4-20h enzymes, which in turn would tri-methylate H4-K20 (Schotta *et al.*, 2004). The role of RNA in this model will be discussed in section 1.3.2.

Recently HP1 dosage was shown to participate in the silencing of mammalian, constitutively expressed, euchromatic genes, that is, genes that are not physically linked to large blocks of heterochromatin (Festenstein *et al.*, 1999). A murine CD2 transgene reporter was shown to variegate in the thymus in response to HP1 dose and

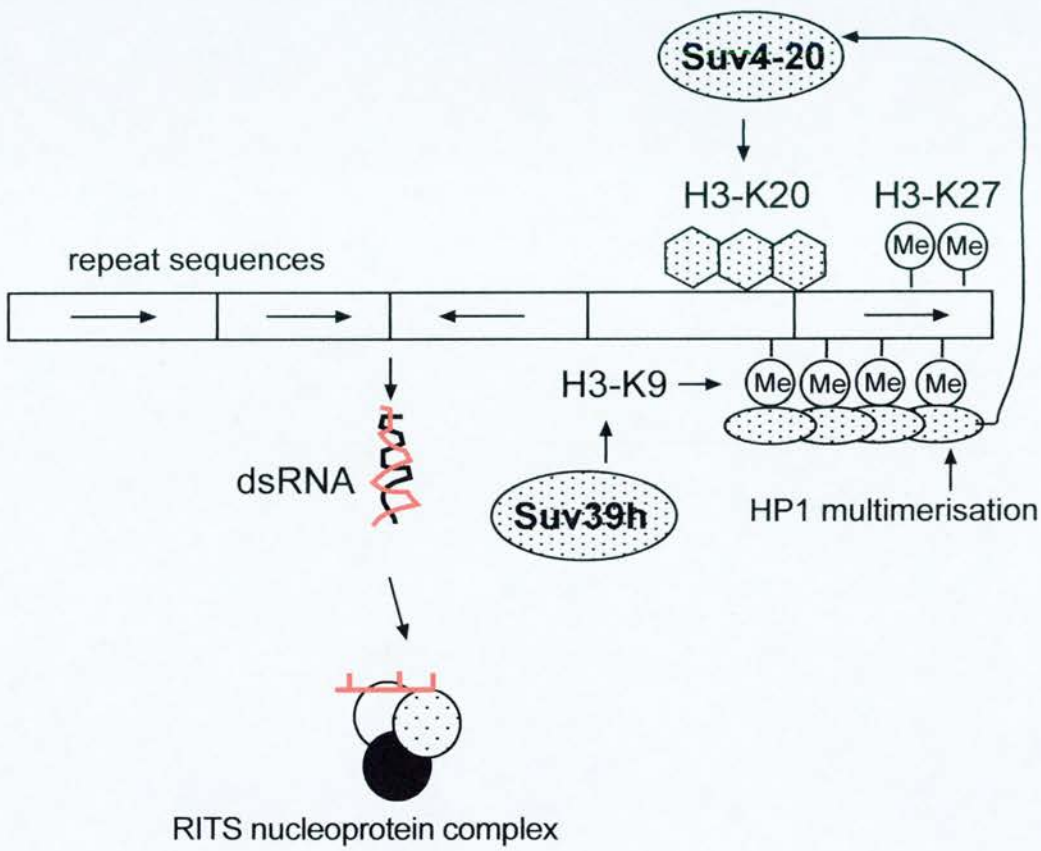


Figure 1.5 Current model of constitutive heterochromatin formation
Heterochromatic tandem repeats give rise to ds RNA that is cleaved by a ds RNA ribonuclease termed Dicer into 21-25 nt siRNAs. In *S. pombe* siRNAs, along with a nucleoprotein complex called RITS (RNAi-induced transcriptional silencing), are recruited to centromeres and guide Suv39h to the region. H3-K9 methylation then recruits HP1/Swi6 deposition, which in turn recruits Suv4-20 to tri-methylate H3-K20 residues (section 1.3.2.1).

variegation was either enhanced or repressed by HP1 dose depending on the integration site. However, the coordination of processes such as histone deacetylation, H3-K9 methylation, HP1 deposition and chromatin compaction in facultative heterochromatin are less well studied.

1.3.2 The role of RNA in transcriptional regulation and heterochromatin formation

Transcription of non-coding RNAs is now known to have a significant role in the epigenetic regulation of gene expression and in the initiation of heterochromatin formation (Richards and Elgin, 2002). Initial clues that an epigenetic regulatory RNA component existed were obtained by the identification of untranslated transcripts of *Xist* and *Tsix* that are involved in the initiation of X-inactivation in female cells by coating of the inactivated X-chromosome (Clemson *et al.*, 1998). Non-coding transcripts appear to be involved in genomic imprinting, for example, *Air* is involved in imprinting at the *Igf2r* cluster (Sleutels *et al.*, 2002)

In 1999, Hamilton and Baulcombe identified short RNAs, 21-25 nucleotides (nt) in length, which were responsible for specifically targeting endogenous plant genes for transcriptional repression. The idea that RNA might interfere with protein expression indicated an RNA interference model (RNAi) where double stranded RNA (ds RNA) molecules could target specific mRNA molecules for degradation (Elbashir *et al.*, 2001). Unexpectedly, the proteins responsible for RNAi (the RNAi machinery) have also been found to have a central role in assembly of silent, condensed heterochromatin (reviewed by Lippman and Martienssen, 2004). Repeat RNA transcripts are processed into short interfering RNAs (siRNAs) that direct chromatin modification at centromeric and telomeric regions in fission yeast, *Tetrahymena*, *Drosophila* and *Arabidopsis* (Hall *et al.*, 2002; Volpe *et al.*, 2002; Taverna *et al.*, 2002; Pal-Bhadra *et al.*, 2002; Zilberman *et al.*, 2003). In yeast, RNAi mutants defective in vital components of the RNAi machinery exhibit reduced H3-K9 methylation and both HP1 and H3-K9 methylation are lost from centromeric reporter genes (Volpe *et al.*, 2002). This ultimately disrupts chromosome segregation during mitosis because HP1 cannot recruit cohesin (Bernard *et al.*,

2001). In *S. pombe*, an RNAi-induced transcriptional gene-silencing complex (RITS) has been purified which contains RNA-binding protein (Ago) and a chromodomain protein (Chp1), which can bind to centromeres (Verdel *et al.*, 2004). There is no evidence for the recruitment of such a complex in mammals. In mammals, very few siRNAs have been cloned, however, pericentromeric localisation of H3-K9 methylation and HP1 is abolished by RNase treatment (Maison *et al.*, 2002). Therefore an RNA signal may provide an entrée in the formation of heterochromatin whereby Suv39h enzymes would be subsequently targeted to repeat-rich sequences and components of the RNAi machinery at heterochromatin (Fig. 1.5).

1.3.3 ATP dependent chromatin remodelling

Nucleosomes can also undergo ATP hydrolysis, and to date there are four classes of nuclear proteins that can remodel chromatin in this way, based on the nature of their ATPase subunits (reviewed by Tsukiyama, 2002).

The first class constitutes the SWI-SNF (mating type switching/sucrose non-fermenting) complex, and is the prototypical chromatin-remodelling complex, originally identified in budding yeast (Stern *et al.*, 1984). This complex has been found to disrupt chromatin structure and facilitate the binding of transcriptional regulators to promoters (Hirschhorn *et al.*, 1992; Krebs *et al.*, 1999) but SWI/SNF is also involved in transcriptional repression (Sudarsanam *et al.*, 2000). SWI/SNF has an extensive repertoire of biochemical activities *in vitro*. It can reposition nucleosomes in an ATP-dependent pattern by sliding histone octamers to other sites on the same DNA molecule (*cis*-displacement) (Whitehouse *et al.*, 1999) or by transfering histone octamers to other DNA molecules (*trans*-displacement) (Lorch *et al.*, 1999). In mammalian cells, there are two characterised Swi2/Snf2-like ATP-ases, BRM and BRG1. Both of these associate with tumour suppressor proteins such as Rb and BRCA1. $BRG^{-/-}$knockout mice also indicate susceptibility to tumour growth and are embryonic lethal (Bultman *et al.*, 2000). Mutations in the human *ATRX* gene (X-linked α-thalassemia/mental retardation syndrome) implicate SWI-SNF remodelling in differentiation and development because ATRX is an ATP-

dependent type II class SNF2 helicase (Picketts *et al.*, 1996). Recently, mass spectrometry and western blotting identified BRM, BRG1 and KAP-1, in addition to HDAC3, as members of N-CoR-1 (nuclear receptor-1) complex (Underhill *et al.*, 2000). This would allow coupling of gene specific regulation by KRAB-ZFPs with chromatin remodelling proteins via KAP-1 and N-CoR-1.

The other class of ATP-dependent chromatin remodelling proteins are the ISWI (imitation switch) family, originally identified in *Drosophila*, the CHD (chromodomain ATPase) and IN080 classes. ISWI homologues in yeast (Isw1 and Isw2) and human (hSNFL and SNF2H) have been identified. Initial studies on Isw2 mutants have shown that this complex is involved in control of meiotic growth genes and the formation of nuclease resistant chromatin structures (Goldmark *et al.*, 2000). Mouse Snf2h is required for early pre-implantation development (Stopka *et al.*, 2003) and recently SNF2H was found to target the Williams syndrome transcription factor (WSTF) to replication foci through direct interaction with PCNA (Poot *et al.*, 2004). The CHD class is generally thought to be involved in transcriptional repression since Mi-2 is a characterised member of this family, and is found in conjunction with HDACs and MBD proteins in the NuRD complex (Zhang *et al.*, 1998; Wade *et al.*, 1999). The PHD and bromodomains of KAP-1 form a co-operative unit that interacts with Mi-2 α (Schultz *et al.*, 2001). This discovery showed that the KRAB-KAP-1 interaction represses transcription by employing the activities of chromatin associated enzymes such as HDACs and NuRD and outlined a mechanism of repression in which the largest transcription factor family in the genome are members.

1.3.4 KRAB-KAP-1 mediated effects on chromatin structure

KAP-1 coordinates all the machinery required for repression of an integrated luciferase transgene, recruited via an inducible KRAB-domain containing protein (Ayyanathen *et al.*, 2003). The KRAB box, is required for localised compaction of the promoter region (assayed by nuclease accessibility assays) and the enrichment of H3-K9 methylation, KAP-1, HP1 and SETDB1 proteins at the promoter (Schultz *et*

al., 2002; Ayyanathen *et al.*, 2003). Repression of KRAB-ZFP target genes is now known to involve a cascade of events starting with the recruitment of KAP-1 to specific genes via the KRAB domain (Fig. 1.6). Deacetylation of histones and ATP-dependent chromatin remodelling then occurs as a consequence of KAP-1 interaction with NuRD, followed by recruitment of SETDB1 for methylation of H3-K9 in euchromatic regions. Interaction and phosphorylation of HP1 proteins by KAP-1 is thought to bring about localised heterochromatinisation of the loci specified by KRAB-ZFPs or the locus itself may be re-located to heterochromatin. Multimerisation and higher order chromatin structure may add further complexity to this model because KAP-1 is a homotrimer *in vivo* and two molecules of HP1 can bind per KAP-1 monomer (Peng *et al.*, 2000; Peng *et al.*, 2002). KAP-1 does not hetero-oligomerise with other members of the TIF1 family. A potential role for KAP-1 modification by sumolyation has been observed but its function is not yet assessed (Seeler *et al.*, 2001).

1.4 Nuclear compartments and their role in chromatin organisation and gene regulation

Another mechanism of epigenetic regulation involves the 3-D organisation of chromatin, and therefore individual genes and their cognitive transcriptional regulators, in the nucleus. Figure 1.7 depicts some of the sub-structures compartmentalised in the mammalian nucleus, many of which are outlined below.

1.4.1 The nuclear periphery and gene silencing

The nuclear periphery of mammalian cells consists of the nuclear membrane, a lipid bi-layer adjacent to the nucleoplasm, and the nuclear lamina, a filamentous network of proteins that is located adjacent to the inner nuclear membrane (INM). It is proposed that the lamina acts to bridge chromatin to the nuclear envelope. The INM contains LEM domain proteins that can bind DNA and chromatin *in vitro* (Dechat *et al.*, 2000). The lamina associated proteins LAP-2β and HA95 also contain LEM domains and are implicated in initiation of transcription (Martins *et al.*, 2003).

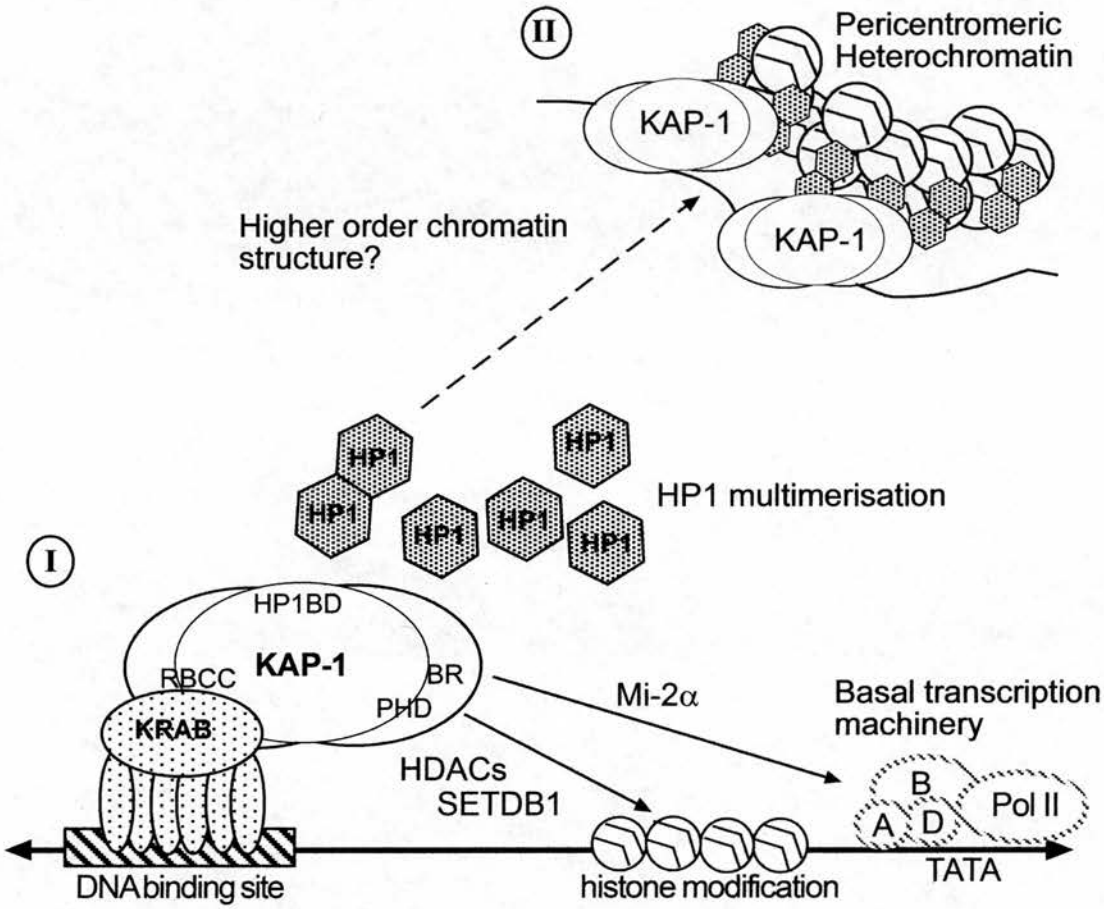


Figure 1.6 Proposed model for KRAB-KAP-1 mediated repression
KRAB-ZFPs target specific genes for repression mediated by the co-repressor KAP-1. KAP-1 recruits histone deacetylases (HDACs) and ATP-dependent chromatin remodeling factors (Mi-2a) to establish repression. SETDB1 can bind to KAP-1 and methylate H3-K9 in euchromatic regions. Interaction of HP1 proteins with KAP-1 and/or methylated H3-K9 may bring about localised heterochromatinisation of the target loci (I) or the loci may be re-located to heterochromatin (II). The kinase activity associated with KAP-1 may also affect gene regulation by interaction with the transcription machinery. *RBCC*, Ring, BB, coiled coli; *PHD*, plant homeodomain; *Br*, bromodomain; *HP1BD*, HP1 binding domain.

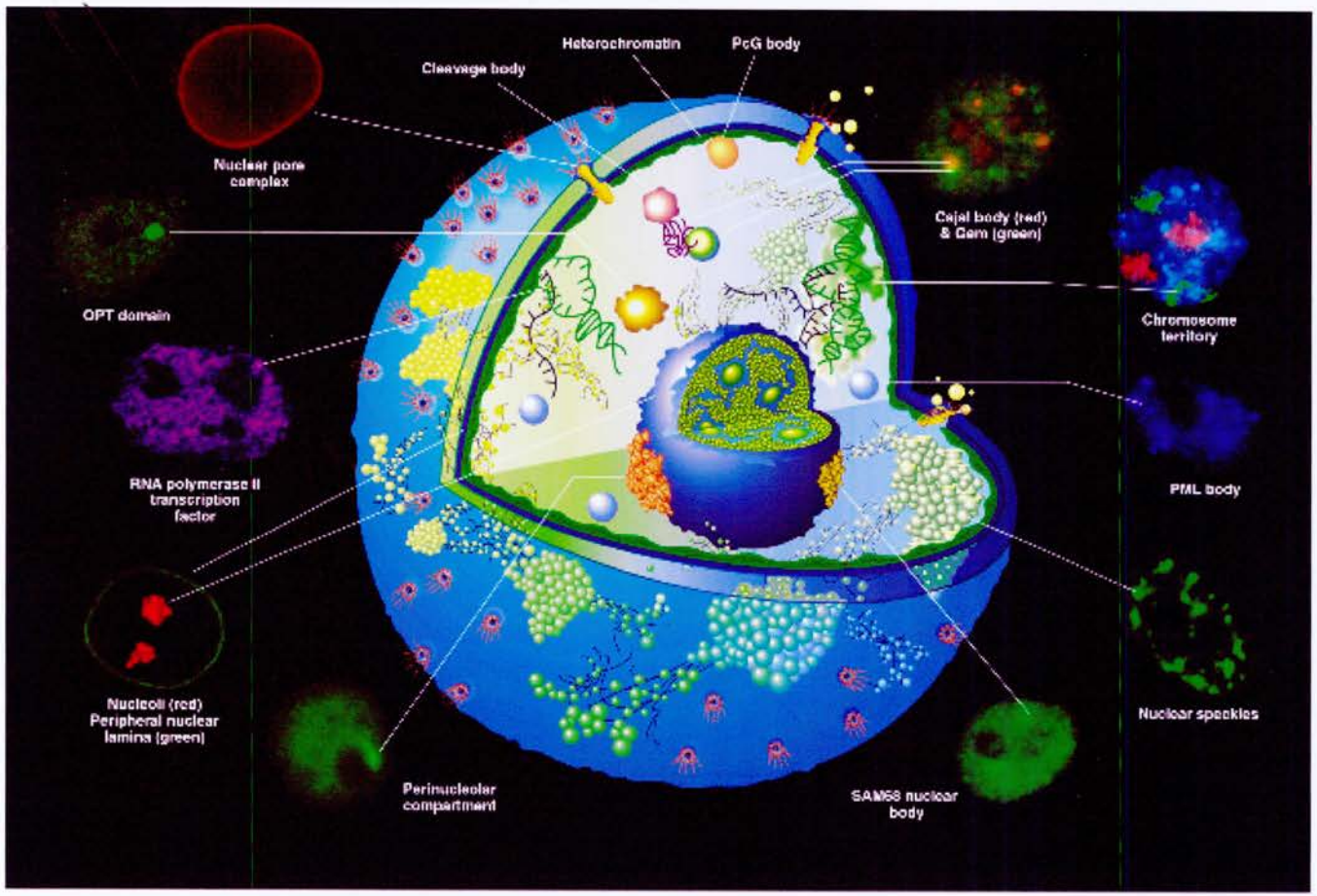


Figure 1.7 Compartmentalisation of the interphase nucleus
Sub-compartments of the interphase nucleus identified by light microscopy. Taken from Spector *et al*, 2001.

Lamins and the INM protein emerin are mutated in the laminopathy disorders, and it has been proposed that a consequent dysregulation of gene expression could underlie some of the disease pathologies (Mounkes *et al.*, 2003).

The nuclear periphery is characteristically associated with gene silencing. In *S. cerevisiae*, heritable inactivation of telomeric genes and the silent mating type locus cluster near the nuclear periphery (Gravel *et al.*, 1998). Sir proteins are maintained at high concentrations in this compartment and the proteins yKu, Mlps and Esc have been shown to function in tethering of silent loci to the nuclear membrane (Gartenberg *et al.*, 2000; Galy *et al.*, 2000; Andrulis *et al.*, 2002). However, there is no evidence that homologues of these proteins play a similar role in mammals. FISH studies suggest that relatively gene poor chromosomes (HSA4, 13 and 18) are situated around the periphery of the nucleus (Croft *et al.*, 1999; Boyle *et al.*, 2001).

1.4.2 The nuclear interior and transcription

Early-replicating DNA is preferentially located in the nuclear interior (Sadoni *et al.*, 1999) consistent with clustering of gene rich chromosomes there (Boyle *et al.*, 2001). Sub-nuclear positioning represents a way of regulating transcription and recombination of endogenous Igκ loci during lymphocyte development (Kosak *et al.*, 2002). FISH shows that Igκ loci are preferentially positioned at the nuclear periphery in haematopoietic progenitors and pro-T cells but are centrally configured in pro-B nuclei. Moreover, in pre-B cells, one allele of Igκ is preferentially packaged into active chromatin characterised by H3-K4 methylation and histone deacetylation, whilst the other allele is recruited to heterochromatin, where it is associated with HP1-γ and Ikaros (Goldmit *et al.*, 2004). Other genes when active have been observed at more internal sites (Chambeyron *et al.*, 2004; Zink *et al.*, 2004).

Arguing against such preferential positioning of transcription, sites of transcription have been visualised throughout the entire nucleus (Szentirmay and Swadogo, 2000). Labelled nascent RNAs are distributed throughout the nucleoplasm; they are not limited to the interior of the nucleus (Verschure *et al.*, 1999). Antibodies against

hyperphosphorylated forms of RNA polymerase II (Zeng *et al.*, 1997), and detection of BrUTP incorporation (Jackson *et al.*, 1993) showed transcription foci to be randomly positioned.

1.4.3 Transcription and the chromosome territory

Initial studies showing active gene transcription on the surface of chromosome territories were based on FISH analyses of single selected loci and they suggested that inactive loci are positioned towards the interior of chromosome territories, while active genes are found on the surface (Volpi *et al.*, 2000). The positions of contiguous genes spread along the length of a segment of HSA11 were mapped with respect to their position within the chromosome territory (Mahy *et al.*, 2002b). In this study, active genes were not always on the territory surface, but were found located within chromosome territories. Inhibition of transcription by Actinomycin-D (Act-D), reduced the percentage of loci visualised outside the territory, suggesting that chromatin looping/movement is influenced but not driven by transcription (Mahy *et al.*, 2002a).

More recently, the co-linearity of genes in the *HoxB* cluster was studied in the context of chromatin structure and nuclear re-organisation upon induction of *HoxB* expression by retinoic acid (Chambeyron *et al.*, 2004). ChIP was used to identify changes in histone modification, and chromatin decondensation was observed upon transcription. However, the progressive looping out of genes from the chromosome territory towards the nuclear centre, is proposed to be necessary for the activation of high levels of transcription since decondensation and histone modification is not restricted to the most 3' gene, *Hoxb1*.

Gene dense regions of HSA11, the MHC (major histocompatibility complex) locus and the EDC (epidermal differentiation complex) locus are excluded from chromosome territories in FISH studies (Mahy *et al.*, 2002a; Volpi *et al.*, 2000; Williams *et al.*, 2002). Moreover, a novel approach combining chromatin sedimentation analysis with FISH was used to identify 'open' chromatin fibres that correlate with regions of high gene density (Gilbert *et al.*, 2004) but not with gene

expression. In this way, inactive genes are found in domains of 'open' chromatin and active genes in regions of 'closed' chromatin. The fact that chromatin fibre structure impacts on levels of decondensation suggests that domains of 'open' chromatin may create a nuclear environment that facilitates transcriptional activation.

1.4.4 Centromeric heterochromatin

When a block of heterochromatin was placed adjacent to the euchromatic *Drosophila Brown* gene, bwD, gene expression was trans-inactivated, along with the normal bwD allele and silencing correlated with relocation of both bwD copies to centromeric heterochromatin (Platero *et al.*, 1998). This suggested that transcriptional repression could be accompanied by relocation of genes to heterochromatic compartments of the nucleus. A similar scenario was reported in mammalian cells where inactive genes were found to co-localise with Ikaros protein and centromeric heterochromatin during lymphocyte proliferation (Brown *et al.*, 1997; Brown *et al.*, 1999). *Rag* and *TdT* alleles reposition to centromeric domains in primary $CD4^{+}8^{+}$ thymocytes (Brown, *et al.*, 1999). Ikaros is found complexed to NuRD and so may function in the remodelling of chromatin, similar to KAP-1 (O'Neil *et al.*, 2000). It has been suggested that Ikaros-binding influences gene expression in B-lymphocytes by the recruitment of chromatin remodelling factors, rather than centromeric localisation *per se* (Sabbattini *et al.*, 2001) as silencing is exerted through a direct effect on the promoter. In the KRAB-KAP-1 system, centromeric localisation of KRAB-ZFPs does not occur when the interaction with KAP-1 is disrupted and this correlates with lack of repression activity (Matsuda *et al.*, 2001).

Topological constraints would prevent all silenced genes within a cell from being localised at pericentromeric heterochromatin, and in some cases the association with heterochromatin is not enough to silence transcription (Nagaki *et al.*, 2004; reviewed by Dillon and Festenstein, 2002). Translocation of proteins and genes may be a consequence of silencing, rather than causative. For example, on recruitment of C/EBPβ (CAAT/enhancer proteins) and –δ to pericentromeric heterochromatin, histone acetylation and TBP binding is potentiated consistent with non-silencing

(Zhang *et al.*, 2001; Schaufele *et al.*, 2001). Interestingly, KAP-1 physically interacts with C/EBPβ and functions as a co-activator in myeloid differentiation (Rooney *et al.*, 2001).

1.4.5 The nucleolus

The most obvious nuclear sub-structure is the nucleolus. It is composed of the fibrillar centre, a dense fibrillar component and the granular component (reviewed by Carmo-Fonseca *et al*, 2000). Transcription and processing of rDNA takes place in the nucleolus, and the nucleolus forms around acrocentric rDNA clusters in humans (Sullivan *et al.*, 2001). The nucleolus is therefore linked to chromosome territory organisation through these acrocentric associations but transcription of rDNA does not appear to direct association of human acrocentric chromosomes to form the nucleolus. The nucleolus is also responsible for pre-ribosomal assembly, sequestration of proteins and cell-cycle control (Carmo-Fonseca, 2000) and the satellite sequences of HSA1 and the Y chromosome are adjacent to the nucleolus during G1 (Bridger *et al.*, 1998).

1.4.6 Positioning relative to other nuclear bodies

Genes and protein complexes specifically co –localise with a number of other nuclear bodies. These include PML nuclear bodies, Cajal (coiled bodies), and nuclear speckles (reviewed by Dundr and Misteli, 2001).

PML nuclear bodies are seen as dense fibrillar ring structures by microscopy. The MHC locus associates with PML bodies (Wang *et al.*, 2004) but the relationship between them is poorly understood. In $PML^{-/-}$ transgenic lines, MHC I protein is down-regulated (Zheng *et al.*, 1998). However, in RNAi knockdown experiments of PML, MHC I expression has been shown to be unaffected (Bruno *et al.*, 2003; Wang *et al.*, 2004). Cajal bodies are also associated with specific gene loci that include the histone gene clusters, U1 and U2 small nuclear RNA genes, and the U3 small nucleolar RNA genes (Frey *et al.*, 1999). β-actin and collagen 1α1, are actively

transcribed genes that also associate with splicing speckles (Xing *et al.*, 1995; Shopland *et al.*, 2003).

Transcriptionally active genes have been proposed to associate with 'transcription factories'. These are discrete nuclear sites of nascent RNA production and concentrated transcriptional components, such as RNA polymerase (Iborra *et al.*, 1996; Jackson *et al.*, 1998; Verschure *et al.*, 1999; Zaidi *et al.*, 2002). Several active genes are dynamically organised into shared, pre-assembled, nuclear sub-compartments/transcription factories, rather than the specific recruitment of these transcription complexes to individual genes (Osborne *et al.*, 2004). It is not yet known whether a similar mechanism of transcriptional control occurs for the repression of euchromatic genes.

1.5 Proposed Research

1.5.1 Background to proposal

Prior to my PhD, a visual gene-trap screen in mouse ES and F9 cells had been performed for the large-scale identification of nuclear proteins and the sub-nuclear compartments in which they reside (Tate *et al.*, 1998; Sutherland *et al.*, 2001). The importance of this kind of study is revealed firstly by the fact that mis-localised nuclear proteins are involved in human genetic disease, cancers and in virally infected cells (Marsh *et al.*, 1998; Koken *et al.*, 1997; Bell *et al.*, 2000; Chen *et al.*, 2004). Secondly, determination of sub-nuclear localisation can give important clues as to the function of novel proteins (McDowell *et al.*, 1999). Also, because common protein signatures are often found in proteins from a given nuclear compartment, this allows for the possible prediction of biological pathways in which novel proteins could function (Bickmore and Sutherland, 2002).

Seven KRAB domain-containing proteins were identified in this gene-trap screen. Prior to 2001, the specific functions of some KRAB-ZFPs, and a few target sequences in humans, had recently become known (Casademunt *et al.*, 1999; Skapek

et al., 2000; Zheng *et al.*, 2000). Apart from this, very little was known about the specific biological functions, and mechanisms of repression, employed by this protein family.

The trapping of these genes therefore provided a way in which to study this protein family that had not previously been utilised. Initial characterisation of some of these trapped proteins revealed that the majority were novel (Sutherland *et al.*, 2001) and that some were found to localise to pericentromeric heterochromatin, in a sub-population of mouse ES cells (Sutherland *et al.*, 2001). This variable localisation pattern of KRAB-ZFPs had only been published by two other groups (Payen *et al.*, 1998 and Matsuda *et al.*, 2001) with the suggestion that there may be a common function at pericentromeric heterochromatin for KRAB-domain containing proteins. The interaction between KAP-1 and HP1 supported this proposal, suggesting a mutual role for KRAB-ZFPs in the sequence-specific targeting of genes to heterochromatin, where they could undergo gene silencing (Ryan *et al.*, 1999; Nielsen *et al.*, 1999). Fuelled by these observations, I believed that the function of KRAB-ZFPs at pericentromeric heterochromatin might be rooted in their common mechanism of repression of target genes but *in vivo* evidence for such a model remained elusive. The gene-trap screen hence paved way for a study, which could involve the use of both functional and descriptive methods, to contribute to a better understanding of this protein family and individual KRAB-ZFPs.

1.5.2 Initial proposal

In light of the above, my project was roughly divided into two parts. Firstly, since knockout mice existed for only one KRAB-ZFP at that time (Casademunt *et al.*, 1999), an initial aim of the project was to establish a mutant (gene-trapped) KRAB-ZFP mouse line, as a first step towards the functional analysis of a specific KRAB-ZFP (Chapter 6). Likewise, in mice, no target genes were known for any KRAB-ZFPs at this time so it was thought that the establishment of mouse lines, as well as giving specific clues as to the function and expression pattern of a particular KRAB-ZFP, would also prove to be useful for target gene assessment. Secondly, the fact that only one study actually describes the localisation of an endogenous KRAB-ZFP

to pericentromeric heterochromatin (Payen *et. al.*, in 1998), and the fact that the majority of the literature describes the localisation of many epitope-tagged KRAB-ZFPs to the cytoplasm, the nucleus and other non-heterochromatic sub-structures within the nucleus (De Lucchini *et al.*, 1991; Grondin *et al.*, 1997; Payen *et al.*, 1998; Huang *et al.*, 1999; Yano *et al.*, 2000; Côté *et al.*, 2001), suggested that more *in vivo* localisation studies were needed. I therefore proposed to make a specific antibody to a gene-trapped KRAB-containing protein to check that gene-trapped versions behave in the same way as their endogenous counterpart (Chapter 4). In addition, no research had yet been published on how, or why, the association of KAP-1 and/or KRAB-ZFPs with centromeric heterochromatin could be induced and how this may in turn result in heterochromatin-mediated gene silencing. Hence, I decided to broach this question by examining how the different stages of the cell cycle and the effects of cell differentiation may affect the sub-nuclear localisation of the gene-trapped KRAB-ZFPs and KAP-1 (Chapter 4).

Chapter 2

Materials and Methods

2.1 Microbiology

All chemicals named throughout this chapter were analytical grade and supplied by Sigma or Roche unless otherwise stated. Technical staff prepared, autoclaved and stored general solutions at the required temperature. Nucleic acid manipulations were performed in 1.5 ml or 0.5 ml microcentrifuge tubes. All centrifugations were carried out at RT unless otherwise stated.

2.1.1 Growth of bacterial strains

Competent *Escherichia coli* (*E. coli*) XL1-blue or DH5-α were grown with vigorous shaking or on inverted L-agar plates at 37°C o/n in Luria-Bertani (LB) medium containing the appropriate antibiotic and used in the preparation of plasmid DNA for all cloning procedures.

2.1.2 Generation of competent bacteria

Competent DH5α cells were purchased from Invitrogen™ (sub-cloning or library efficient). Competent cells for all other strains were prepared as follows: A 300 ml *E. coli* culture was grown to an optical density of approximately 0.3 at 550 nm with inoculation from an o/n 5 ml starter culture. After pelleting the cells at 6000 *g* in a Sorval GSA rotor for 10 min at 4°C, they were resuspended in 5 ml 10 mM $CaCl_2$ and incubated on ice for 15 min. Cells were pelleted once more and resuspended in 10 ml freezing mix (15% glycerol [v/v], 75 mM $CaCl_2$, 10 mM KCl, 10 mM MOPs (3-[N-Morpholino] propanesulfonic acid; 4-Morpholinopropanesulfonic acid, pH 7). 300 µl aliquots were snap frozen on dry ice and stored at -80°C until required.

2.1.3 Bacterial transformations

Competent cells were incubated on ice with ~10-100 ng DNA for 30 min. Heat shock treatment at 42°C for 45 sec enabled bacterial transformation. Bacteria where then grown with shaking at 37°C in 500 µl of LB medium for at least 45 min to recover, before being plated on L-agar plates containing the selection antibiotic

(ampicillin [50 μg/ml], kanamycin [50 μg/ml], carbenicillin [50 μg/ml] or chloramphenicol [34 μg/ml]). Plates with transformed bacteria were grown o/n at 37°C.

2.2 Preparation and manipulation of DNA

2.2.1 Plasmid DNA isolation

Single bacterial colonies were picked from agar plates and used to inoculate 5 ml or 100 ml of LB medium with antibiotic for o/n growth at 37°C. Plasmid DNA was then extracted in one of two ways. Either commercial mini- or maxiprep kits (QIAGEN® Plasmid Mini/Maxi Kits) were used according to the manufacturer's instructions; or when clean, supercoiled plasmid DNA was required, by alkaline lysis of bacteria and isolation of the DNA on caesium chloride (CsCl) gradients as described by Sambrook *et al*, (2001). Briefly, bacteria were incubated in resuspension buffer (50 mM Tris-HCl [pH 8], 10 mM EDTA [pH 8], 100 μg/ml RNase A [ribonuclease A]) before lysis under alkaline conditions (0.2 M NaOH, 1% SDS [sodium dodecyl sulfate] [w/v]). Chromosomal DNA was precipitated by addition of 3 M potassium acetate and 2 M acetic acid, pH 5.2. After centrifugation at 4°C for 15 min at 20,000 *g*, the supernatant was passed through muslin. 0.6 vol of propan-2-ol was added to precipitate the DNA in the supernatant before pelleting and drying. The pellet was resuspended in TE (Tris-EDTA) buffer (10 mM Tris-HCl, 1 mM EDTA [ethylenediaminotetra-acetic acid], pH 8) and 1 *g* of CsCl was allowed to dissolve per ml of solution before addition of 10 mg/ml ethidium bromide (EtBr, 2,7-diamino-10-ethyl-9-phenyl-phenathridium bromide). After centrifugation at 12,000 *g*, suspended DNA was loaded into 2.4 ml tubes (Beckman) and spun at 16°C in a Beckman Ti100.3 rotor at 80,000 rpm overnight (acceleration programme 2, deceleration programme 9). Supercoiled DNA was obtained by puncturing the side of the tube with a 21 gauge needle and drawing out the appropriate DNA band (Sambrook *et al*., 2001). EtBr is removed from the DNA by repetitive extractions with equal volumes of water-saturated butan-1-ol. DNA was precipitated with 3 vol

70% ethanol (v/v with dH_2O) before resuspension in TE (typical yield, ~1-2 mg/ml plasmid DNA).

2.2.2 Phenol/chloroform extractions of DNA

An equal volume of phenol/chloroform (50% buffered phenol/48% chloroform [v/v], 0.5% 3-methyl-1-butanol [v/v], pH > 7.8) was added to the DNA preparation and vortexed. This was microcentrifuged at 12,000 *g* for 10 min. The top aqueous layer was decanted into a fresh tube and the chloroform step repeated if the DNA was to be used for cloning. Ethanol precipitation was then performed to concentrate the DNA (section 2.2.3).

2.2.3 Ethanol precipitation

0.1 vol 3 M sodium acetate, pH 5.2 and 2 vol 100% ice cold ethanol were added to the DNA sample. After vortexing, the sample was microcentrifuged at 12,000 *g* for 10 min. The supernatant was discarded and the pellet washed in 70% ethanol. Centrifugation was repeated and the supernatant discarded. Once the pellet had dried at RT, it was resuspended in an appropriate volume of dH_2O.

2.2.4 Mouse genomic DNA extraction

Tissue samples (mainly ~2 mm mouse tail biopsies) were allowed to dissolve in 600 µl of DNA lysis buffer (100 mM EDTA [pH 8], 50 mM Tris-HCl [pH 8], 100 mM NaCl, 1% SDS [w/v]). Protein was digested by addition of 30µl 10 mg/ml proteinase K (Roche) and incubation at 55°C o/n. The DNA was extracted by addition of 420 µl phenol/chloroform and gentle mixing on a rotor for at least 2 hr. Samples were microcentrifuged at 12,000 *g* for 10 min. Approximately 500 µl of the top aqueous layer was decanted to a fresh tube and the DNA precipitated out of this layer upon addition of 500 µl of propan-2-ol (no additional salt required since DNA lysis buffer has an adequate amount). After gentle inversion of tubes, blunt ended glass rods were used to collect the precipitated DNA for drying at RT. Genomic DNA was resuspended in an appropriate volume of TE buffer (~300 µl for a tail biopsy. Typical yield, ~ 300 µg/µl genomic DNA) and 1 µl and 20 µl of this solution were

used for PCR genotyping reactions (section 2.2.12) and for Southern blotting (section 2.2.14) respectively.

2.2.5 DNA digestion

DNA was digested with restriction enzymes (New England Biolabs, NEB) in the required reaction buffer according to the manufacturer's instructions. Enzyme was added at a concentration of 5-10 U/μg DNA with the total volume of the enzyme never exceeding 10% of the total reaction volume so as to prevent the inhibition of digestion by glycerol. Double digests were performed where enzyme buffer requirements and temperature were compatible. When incompatible, they were performed sequentially after a phenol/chloroform extraction and ethanol precipitation (section 2.2.2 and 2.2.3).

2.2.6 Filling in 3' recessed ends of DNA

When cohesive ends needed to be made blunt for subsequent ligations (section 2.2.11), the 3' to 5' exonuclease activity of Klenow fragment (Roche) at a concentration of 0.5 U/μg DNA template was utilised to fill in 3' recessed ends with 100 μM of the relevant dNTPs (deoxynucleotide triphosphates) (Abgene).

2.2.7 Agarose gel electrophoresis

1-2% horizontal agarose ('HiPure' Low EEO agarose, Biogene UK) gels (w/v) were used in conjunction with 1x TAE (90 mM Tris-HCl [pH 8], 90 mM glacial acetic acid, 2 mM EDTA [pH 8]) electrophoresis buffer to resolve DNA samples. EtBr was added to the molten agarose to a final concentration of 0.5 μg/ml before pouring to allow visualisation of the DNA under ultraviolet light (UV). 6x loading buffer (15% Ficoll™400 [Amersham Biosciences], 0.25% Orange G [Sigma] both w/v in dH_2O) stored at RT was added to all DNA samples before loading. 500 ng of the DNA size markers ϕX174 DNA/*HaeIII* (Promega) or λ DNA/*HindIII* (Invitrogen™) were loaded into a well on each gel to enable the sizing and quantification of DNA fragments. Stained DNA was photographed using the EagleEye II system (Stratagene®) and a thermal printer (Mitsubushi).

2.2.8 Gel purification of DNA fragments

Clean razor blades were used to dissect agarose gel slices containing the required DNA fragments and a commercial kit (Hybaid Recovery™ DNA Purification Kit II, Hybaid or QIAquick® Gel Extraction Kit, Oiagen) was used to extract this DNA from the slices. In brief, gel slices are melted in equilibration buffer and this solution is then passed over a silica-gel matrix that selectively binds DNA. Washing this membrane removes EtBr and residual agarose and the DNA is eluted into the required volume of dH_2O.

2.2.9 Analysis of DNA quantity and quality

Quantification of DNA was measured either by comparison of the unknown quantity of DNA with a known quantity on an agarose gel, or by measuring its absorbance in a spectrophotometer (Ultrospec 3000pro, Amersham Biosciences) at 260 nm (A_{260}) (Sambrook *et al.*, 2001).

2.2.10 Dephosphorylation of DNA fragments

Prior to ligation, vector DNA was treated with calf intestinal alkaline phosphatase (CIP) (Roche) in accordance with manufacturer's instructions to dephosphorylate the 5' termini of vector DNA. This prevents self ligation of the empty plasmid DNA consequently increasing the yield of positive ligation products since successful ligation should only be achieved when phosphates are donated to the 5' plasmid ends by untreated DNA fragments with compatible digested ends. The reaction was terminated by phenol/chloroform extraction and ethanol precipitation of DNA (section 2.2.2).

2.2.11 Ligation of DNA fragments

Fragments with cohesive complementary ends were ligated o/n at 16°C, using 1 U of T4 DNA ligase (NEB), 1% NP-40 (Nonidet P-40) (v/v) and the appropriate volume of 10x reaction buffer in a 20 µl ligation reaction. Ligations were usually set up with a 1:3 and 1:10 vector to insert molar ratio using 20-50 ng of both digested vector and

DNA insert. 10 µl of ligation product was subsequently used for bacterial transformation (section 2.1.3).

2.2.12 Polymerase Chain Reaction

2.2.12.1 Reagents

2.2.12.1.1 dNTPs

dNTPs (Abgene) were purchased as 100 mM stocks of each of the dNTPs. Working stock concentrations of 2.5 mM and 10 mM were prepared for use in genotyping and non-genotyping polymerase chain reactions (PCRs) respectively and used at final concentrations of 0.2 mM (genotyping) and 0.5 mM (non-genotyping). Stocks were stored at -20°C.

2.2.12.1.2 Oligonucleotide primers

PCR primers were designed by selecting a sequence between 17 and 25 nucleotides in length for which the [G+C]:[A+T] ratio was approximately equal. OligoAnalyser (v. 3) at http://www.idtdna.com was used to analyse these primers to reduce the likelihood of choosing primer sets that anneal with each other or self anneal. Primers were purchased from MWG Biotech and Invitrogen™ as lyophilised desalted compounds and a stock concentration of 1 µg/µl made with dH_2O. Working stock concentrations were prepared at 25 ng/µl. Oligonucleotides (oligos) were used at final concentrations of 1 ng/µl and 0.5 ng/µl for genotyping and non-genotyping PCRs, respectively.

2.2.12.1.3 Additional PCR reagents

Plasmid DNA to be amplified during non-genotyping PCRs was used at a concentration of 10 ng/µl. For non-genotyping PCRs, Ampli*Taq*® DNA polymerase (Roche, 5 U/µl) was used at 0.5 µl per 50 µl reaction or when high fidelity PCR products were required (e.g. for subsequent cloning applications), Pfu*Turbo*® DNA polymerase (Stratagene®, 2.5 U/µl) was used at the same concentration as above. For amplification of genomic DNA template (section 2.2.4), Ampli*Taq*® DNA

polymerase was used at 0.2 µl per 25 µl reactions. 10x PCR buffer and 2.5 mM $MgCl_2$, supplied with their respective polymerase, were used at 1:10 and at 1.5 mM respectively. Reaction mixes were made up to 25, 50 or 100 µl total vol with dH_2O.

2.3.12.2 PCR amplification programmes

PCRs were performed in 0.5 ml centrifuge tubes or 96 well PCR plates in a MJ Research DNA Engine Tetrad. The formula below represents a PCR thermal cycle that includes an initial *v* min DNA denaturation step at 94°C followed by 30 cycles of a *w* min denaturation step, a *x* sec annealing step at a temperature specific to the primers in use (usually between 50-60°C), a *y* min elongation step at 72°C, and a final elongation step for *z* min to complete the cycle:

1x(94[*v*m]) 30x(94[*w*m]50[*x*s]72[*y*m]) 1x(72[*z*m])

Formulas resembling the above will be used to describe PCR thermal cycles throughout this thesis. Primer sequences and amplification programmes used in genotyping and non-genotyping PCRs are tabulated below.

Table 2.1 Primers and amplification programmes used for genotyping PCRs

LacZ primers (section 6.2.2)	5' *LacZ* forward	GTTGCGCAGCCTGAATGGCG
	3' *LacZ* reverse	GCCGTCACTCCAACGCAGCA
	Programme	1x(92[3m])32x(92[30s]60[45s]72[45s]) 1x(72[5m])
492 wt primers for PCR (section 6.3.2.1.1) and for Southern/northern probe (section 6.3.2.1.2)	5' wt 492 forward	CACAGACTTGAATGTGC
	3' wt 492 reverse	AGATGGGAGCTCCGACCAAAG
	Programme	1x(92[3m])32x(92[30s]52[45s]72[45s]) 1x(72[5m])

Table 2.2 Primers and amplification programmes used for non-genotyping PCRs

492 linker region for antibody production (section 4.8.1)	5' 492 link forward	ACTGAATTCGGAAGTGACCACTCAG AATGC
	3' 492 link reverse	TGAGCGGCCGCATAGGGTCTCTCAA CAGTGGG
	Programme	1x(94[3m])25x(94[30s]55[1m]72[2.5m])1x (72[5m])
pEGFP-C1/492K (section 4.7)	5' 492p1C1	ACTAGATCTGATACCAAGATGGCAG CAGCC
	3' 492p1C1	AGATGGGAGCTCCGACCAAAG
	Programme	1x(94[3m])25x(94[30s]63[1m]72[2.5m])1x (72[5m])
pEGFP-C1/492ZF (section 4.7)	5'492p2N1/C1	CTTTGGTCGGAGCTCCCATCT
	3' 492p2C1	AGTAAGCTTTCACAGCTTCTTGTGTA CTTTC
	Programme	1x(94[3m])25x(94[30s]63[1m]72[2.5m])1x (72[5m])
pEGFP-N1/492K (section 4.7)	5' 492p1N1	ACTAGATCTCACAGGATCGCCTGATT TCTG
	3' 492p1N1	AGATGGGAGCTCCGACCAAAG
	Programme	1x(94[3m])30x(94[30s]52[1m]72[2.5m]) 1x(72[10m])
pEGFP-N1/492ZF (section 4.7)	5'492p2N1/C1	CTTTGGTCGGAGCTCCCATCT
	3' 492p2N1	AGTAAGCTTCACAGCTTCTTGTGTAC TTTCTG
	Programme	1x(94[3m])30x(94[30s]52[1m]72[2.5m]) 1x(72[10m])

All cloning products were sequenced in order to monitor the integrity of recombinant plasmids and inserts. BigDye Terminator v. 3.1 (ABI Prism) was used as follows: 200 ng of plasmid DNA or PCR product, 0.8 μM sequencing primer and 4 μl BigDye were diluted in dH_2O in accordance with the manufacturer's instructions to a final volume of 10 μl. Reactions were run on a DNA Engine Tetrad (MJ Research) for 25x(96[30s]50[15s]60[4m]. Reactions were cleaned by addition of 0.2 vol 3 M sodium acetate (pH 5.2) and 5x vol 95% ethanol, incubated on ice for 10 min and pelleted at 12,000 *g*, 4°C for 30 min. Pellets were washed once in 70% ethanol and centrifuged as above for 15 min before being left to air dry. Resuspended reactions were run through 50 cm capillaries on POP6 polmer gels (3100 Genetic Analyser, Applied Biosystems) and data collection was performed using v. 3.7 of the supplied software.

2.2.14 Analysis of genomic DNA by Southern blotting and hybridisation

2.2.14.1 The Southern Blotting protocol

Up to 10 μg of genomic DNA was digested in 50 μl reactions o/n using 20-30 U of restriction enzyme in the buffers supplied and under the conditions specified by the manufacturer. The digested DNA was separated by electrophoresis in 0.8% agarose gels containing EtBr (section 2.2.7). The DNA was depurinated by UV nicking in a GS Gene Linker™ UV Chamber (Bio-Rad) and capillary blotted to Hybond-N^+ membrane (Amersham Biosciences) using 0.4 M NaOH as the transfer solution (Reed and Mann, 1985). The membrane was then briefly neutralised at RT in neutralisation solution (0.2 M Tris-HCl [pH 7.4], 2x SSC made from a 20x stock [3 M NaCl, 0.3 M tri-sodium citrate, pH7.4]) and rinsed in 2x SSC for 5 min at RT. Membranes were firstly prehybridised in 20 ml of hybridisation solution (5x Denhardt's reagent made from 50x stock [1% BSA (bovine serum albumin) (w/v), 1% Ficoll®400 (v/v), 1% polyvinylpyrrolidone (w/v)], 0.5% SDS [w/v], 6x SSC), supplemented with 100 μg/ml of fragmented salmon sperm DNA (Sigma) in

cylindrical glass bottles (Amersham Biosciences) at 65°C with continuous rotation. Membranes were prehybridised for 5 hr before addition of freshly boiled probe DNA (section 2.2.14.2) to the prehybridisation solution. After an o/n hybridisation with the probe, membranes were washed in 0.2% SDS (w/v) in 2x SSC at RT for 15 min, then in 0.5% SDS (w/v) in 0.1x SSC at 37°C and 65°C for 1 hr each, rinsed in PBS and allowed to air dry. Membranes were exposed to a storage phosphor screen (BSA cassette, Fuji Film) for 1-3 days and visualised using Fuji Film FLA-5100 v. 2 phosphor imager hardware and Aida Image Analyser v. 3.44 software (both Raytek Scientific Ltd)

2.2.14.2 Probe preparation and radio-labelling of probes

DNA fragments to be radio-labelled for use as radio-active probes in Southern and northern blot analyses were prepared by PCR amplification of cDNA. 25 ng of purified DNA was made up to 11 µl with dH_2O and denatured by boiling for at least 10 min and immediately placed on ice. Denatured DNA was labelled using random priming with the commercial kit, High Prime (Roche) and [α-^{32}P] dCTP. Briefly, 4 µl of High Prime reagent was added to the probe along with 5 µl (50 µCi) [α-^{32}P] dCTP and mixed to a total volume of 20 µl. The reaction mix was incubated at 37°C for 1 to 5 hr, after which it was placed on a G-50 Sephadex Quick Spin column (Roche) and eluted with 80 µl TE (pH 8) to separate the labelled DNA fragments from unincorporated nucleotides. Labelled probe was then denatured by boiling for 10 min, before being added to prehybridised membranes in hybridisation solution.

2.3 Preparation and manipulation of RNA

2.3.1 RNA isolation and purification

A commercial kit (Bio/RNA-Xcell™ from Bio/Gene) with a single step method for the isolation and purification of total RNA from tissues and cells was used according to the manufacturer's instructions. Cells were homogenised in the Bio/RNA-Xcell™ solution by passage firstly through a 21 gauge needle then subsequently a 23 gauge needle, before chloroform addition and centrifugation of the homogenate in a

microcentrifuge at 12,000 *g*. The RNA found in the upper aqueous phase of the homogenate is then mixed with propan-2-ol and an RNA binding resin provided in the kit. The RNA bound resin is pelleted by microcentrifugation and washed in 75% ethanol twice before being eluted in dH_20. A final round of centrifugation allowed the RNA in solution to be eluted from the resin. RNA was quantified by measuring the A_{260} and calculated as described in Sambrook *et al.* (2001).

2.3.2 5' Rapid amplification of cDNA ends (5' RACE)

5' RACE products were generated as previously described (Tate *et al.*, 1998 and Sutherland *et al.*, 2001) using primers purchased from MWG Biotech and Invitrogen™. The sequences of primers described in this section are in table 2.3. RNA is reverse transcribed into first strand cDNA by annealing 10 ng P456 primer specific for *LacZ* to 5 µg total RNA at 70°C (see Fig. 3.1B). Incubation with 10 mM DTT (dithiothreitol), 10 mM dNTPs and 200 U of Superscript II reverse transcriptase (Invitrogen™) at 37°C for 1 hr in a total volume of 20 µl completes first strand synthesis. RNA is hydrolysed at 65°C for 20 min by addition of freshly prepared NaOH to 0.1 M before neutralising with 0.1 M HCl. DNA was microdialysed for 4 hr on 0.025 µm nitrocellulose discs against TE (pH 8) to remove dNTPs and exchange buffer. The volume was made back up to 20 µl with dH_2O. A poly A tail was added to first strand cDNA by incubation at 37°C for 10 min with 2 mM dATP and 30 U of recombinant terminal deoxynucleotidyl transferase (TdT) (Invitrogen™). TdT can catalyse the template-independent addition of dNTPs to the 3' hydroxyl terminus of single stranded (ss) DNA. This tailed DNA was used as a template to synthesise second strand cDNA using primer 56 (Fig. 3.1B). Two consecutive rounds of PCR reaction with nested *LacZ* primers (59, 80, and 79) followed by microdialysis steps on 0.1 µm nitrocellulose disks to size select larger PCR products, were used to amplify the specific 5' DNA sequences of genes. These RACE products were directly sequenced with a –40 *LacZ* sequencing primer (section 2.2.13).

Table 2.3 Primers used for 5' RACE

Primer no.	*Sequence*
P456	5'-CCGTGCATCTGCCAGTTTGAGGGGA
56	5'-GGTTGTGAGCTCTTCTAGATGGT$_{17}$
80	5'-AGTATCGGCCTCAGGAAGATCG
59	5'-GGTTGTGAGCTCTTCTAGATGG
79	5'-ATTCAGGCTGCGCAACTGTTGG
-40 *LacZ* sequencing primer	5'-GTTTTCCCAGTCACGAC

2.3.3 Analysis of RNA by northern blotting

A mouse Multiple Tissue Northern (MTN®) Blot kit was purchased from BD Biosciences (# 7762-1) and used according to the manufacturer's instructions. Briefly, ExpressHyb Solution was heated to 68°C and stirred to dissolve any precipitate. Membranes were prehybridised in 10 ml of ExpressHyb Solution in cylindrical glass bottles (Amersham Biosciences) with continuous rotation for at least 30 min at 68°C. Freshly boiled, radio labelled probe was prepared as described in section 2.2.14.2, added to the prehybridisation mix and left to incubate with continuous rotation at 68°C o/n. The blot was rinsed the following day in 0.05% SDS (w/v) in 2x SSC for 40 min at RT with several changes of the wash solution. The blot was then washed twice in 0.1% SDS (w/v) in 0.1x SSC with continuous rotation for 40 min each at 50°C. Excess wash solution wash removed and the blot immediately covered in plastic wrap and exposed using a storage phosphor screen as previously described (section 2.2.14.1).

Membranes were reused after stripping previous probe from the blot as described below. Sterile dH_2O containing 0.5% SDS was heated to ~90-100°C. The membrane was placed in the heated solution for 10 min. The dH_2O was allowed to cool for 10 min before removing the blot and allowing it to either air-dry for long term storage, or placed in 2x SSC for a subsequent hybridisation.

2.4 Protein preparation and analysis

2.4.1 Total, soluble and insoluble protein extractions from bacterial cells

A 20~40 ml culture (OD_{600}=0.75) containing appropriate antibiotic was induced with 1 mM IPTG (Isopropyl bD-thiogalactopyranoside) (Melford Laboratories Ltd). Cultures were left to grow for 3~4 hr until OD_{600} ~1.6. 1 ml aliquots of cultures were removed before induction and after induction at 1 hr intervals. The pellet was centrifuged for 1 min at 12,000 *g* and resuspended in an appropriate volume of lysis solution (50 mM Tri-HCl [pH 8], 2 mM EDTA [pH 8]) as determined by the OD_{600} reading of the sample. (It is known that when ~160 µl of lysis solution is added to a culture of OD_{600} = 2, protein can be easily visualised by SDS-PAGE [SDS-polyacrylamide gel electrophoresis][2.4.4]). 1 µl of 10 mg/ml lysozyme and 1 µl 1% NP-40 (v/v) were added and the sample was incubated at 37°C for 15 min. 1 µl of 1M $MgSO_4$ and 1 µl of DNase at 1 U/µl (Invitrogen™, 10 U/µl) was then added and incubated as above. The sample was centrifuged at 12,000 *g* for 2 min and the supernatant containing soluble protein removed to a fresh tube. An equal volume of 2x SDS protein loading buffer (125 mM Tris [pH 6.5], 4% SDS [w/v], 10% 2-β-mercaptoethanol [β-ME] [v/v], 20% glycerol [v/v], 0.1% bromophenol blue [w/v]) was added to the supernatant. Boiling the sample for 5 min and sonication on ice for 15 sec at 10 µm produces the soluble protein fraction. Resuspension of the pellet in an appropriate volume of lysis solution as before, with addition of 2x SDS protein loading buffer, boiling and sonication results in the insoluble protein fraction. A total cell protein sample (TPS) was obtained by resuspension of the unlysed cell pellet in an appropriate volume of PBS (phosphate buffered saline) and 2x SDS protein loading buffer, which was boiled and sonicated as above.

2.4.2 Total cell protein extracts from mammalian cells

Cells grown in T25 cm^2 culture flasks were rinsed in PBS, lysed in 300 µl of PBS and 300µl of 2x SDS protein loading buffer, and scraped from the surface of the

flasks. Total protein samples were boiled for 5 min and sonicated on ice at 5 μA for 10 sec before being stored at -20°C until required.

2.4.3 Nuclear protein extracts from mammalian cells

Cells were grown in T75 cm^2 culture flasks until confluent, trypsinized and harvested by centrifugation at 1000 *g* for 3 min. Cells were then washed in PBS and resuspended in 2.5 ml of NBA (Nuclei Extraction Buffer A, 5.5% sucrose [w/v in dH_2O], 10 mM Tris-HCl [pH 8], 85 mM KCl, 0.5 mM spermidine, 250 μM PMSF [phenyl methyl sulfonyl fluoride], 0.2 mM EDTA). 2.5 ml NBB (Nuclei Extraction Buffer B, NBA and 0.1% NP-40 [v/v]) was next added and incubated on ice for 3 min. Nuclei were centrifuged at 500 *g* for 3 min at 4°C, washed in 5 ml of NBA and recentrifuged before resuspension in 200 μl NBA.

To measure the concentration of suspended nuclei, the nuclei were incubated with 0.1 U/μl of DNase I and A_{260} reading taken. Since the concentration of DNA in the pellet equals the concentration of nuclei in the pellet, this OD reading can be used to prepare a nuclear protein extract of known concentration i.e. 1 mg/ml. An equal volume of 2x SDS protein loading buffer was added to the sample and boiled before loading 10 μl (5 μg) on an SDS-PAGE gel (section 2.4.4).

2.4.4 Resolution of proteins by SDS-PAGE

Cell or nuclear protein extracts were harvested and prepared as above (section 2.4.2) and 10 or 15 μl loaded for analysis. Resolution was achieved on denaturing polyacrylamide mini gels (10-12% acrylamide [v/v], 0.39 M Tris-HCl [pH 8.8], 0.1% SDS [w/v], 0.1% ammonium persulfate [w/v], 0.04% TEMED [N, N, N', N-tetramethylethylene diamine] [v/v] in dH_2O) with stacking gels (5% acrylamide [v/v], 0.13 M Tris-HCl [pH 6.8], 0.1% SDS [w/v], 0.1% ammonium persulfate [w/v], 1% TEMED [v/v] in dH_2O) using 30% or 40% acrylamide (29:1 and 37:1 acrylamide: bis-acrylamide [v/v] respectively) (Severn Biotech). Gels were run in electrophoresis tanks (Mighty Small, Hoefer) in Tris-glycine running buffer (25 mM

Tris base, 250 mM glycine, [pH 8.3], 0.1% SDS [w/v]) at 110 V for 2 hr. Pre-stained protein standards (Bio-Rad) were loaded along side samples to aid with analysis.

2.4.5 Visualisation of cellular proteins

After SDS-PAGE, stacking gels were removed and the remaining gel was rinsed in dH_2O. Gels were submerged in coomassie stain (0.25% coomassie brilliant blue R-250 [w/v], 45% methanol [v/v], 10% glacial acetic acid [v/v] in dH_2O) for 1 hr with gentle agitation. The stain was discarded and the gels were incubated o/n in destain (30% methanol, 10% glacial acetic acid in dH_20).

2.4.6 Western Blotting

After SDS-PAGE (section 2.4.4), the stacking gel was removed and protein samples were transferred to polyvinylidene difluoride (PVDF) membrane (Hybond-P, Amersham Biosciences) for further analysis by wet western blotting. In brief, PVDF membrane was soaked in methanol for 10 sec, rinsed in dH_2O then equilibrated in protein transfer buffer (20% methanol [v/v], 24 mM Tris base, 192 mM glycine [pH >8]). Transfer apparatus (GENIE™ blotter, Idea Scientific) was assembled according to the manufacturer's instructions and run for 1 hr at 12 V with the gel and membrane sandwiched by 4 pieces of 3MM paper (Whatman) equilibrated in transfer buffer.

Membranes were then blocked for either 1 hr with agitation or o/n without agitation, in 1% western blocking reagent [v/v] (Roche) in 1x TBS (150 mM NaCl, 10 mM Tris-HCl [pH 7.5]). Primary antibodies (Table 2.4) were diluted in 0.5% western blotting reagent (v/v in 1x TBS) and incubated with the membranes for 1 hr at RT with rotation. Membranes were washed for 3 x 5 min in 1x TBST (1x TBS, 0.05% Tween-20 [v/v]) and incubated with the appropriate secondary antibody (Table 2.5) for 1 hr. Membranes were rewashed for 2x 10 min in 1x TBST, before detection by chemiluminescence (SuperSignal® West Pico Chemiluminescent, Pierce). Signals were exposed on Kodak Biomax film.

Table 2.4 Primary antibody dilutions for western Blots

Antibody	*Species*	*Source*	*Dilution Factor*
Anti-KAP-1 (Ab3)	Rabbit polyclonal	David Schultz (Ryan *et al.*, 1999)	1:1000
Anti-NT-2	Rabbit polyclonal	Yoshihiko Yamada (Tanaka *et al.*, 2002)	1:1250
Anti-492Ab IgG	Rabbit polyclonal	This thesis	1:1500
Anti-492Ab AP	Rabbit polyclonal	"	1:1000
Anti-HP1α (MAB3584)	Mouse monoclonal	Chemicon Int.	1:2000
Anti-HP1β (MAB3448)	Mouse monoclonal	Chemicon Int.	1:2000
Anti-HP1γ (MAB3450)	Mouse monoclonal	Chemicon Int.	1:2000
Anti-fibrillarin Nop1p (mAb D77)	Mouse monoclonal	Gift, J. Aris, (Florida, USA)	1:5000

Table 2.5 Secondary antibody dilutions for western Blots

Antibody	*Species*	*Source*	*Dilution Factor*
Anti-mouse IgG (Fab) Horseradish peroxidase conjugate (HRP)	Goat	Sigma, A9917	1:5000
Anti-rabbit IgG (whole molecule) HRP	Goat	Sigma, A0545	1:5000
Anti-sheep IgG (whole molecule) HRP	Donkey	Sigma, A-3415	1:5000

2.5 Generation of Antibodies

2.5.1 Glutathione-S-transferase purification of fusion proteins

DNA sequence of the antigen was ligated into the expression vector, pGex-4T-1 (Amersham Biosciences) in frame with the glutathione-S-transferase (GST) coding sequence. The expression vector was transformed into *E. coli* BL21-CodonPlus®(DE3)RP cells (Stratagene®) to allow purification of the fusion protein on a glutathione-agarose column. This *E. coli* strain contains the T7 RNA polymerase under control of the lacUV5 promoter and was used for protein over-expression by induction with the galactose analogue IPTG. This strain is designed for high-level expression of mammalian proteins since the cells are protease deficient and contain extra copies of the rare *E. coli* tRNA genes. Cells were grown in chloramphenicol (34 μg/ml) containing media and the fusion protein was purified as follows:

Small-scale cultures were used to determine at what temperature and length of induction maximal soluble fusion protein was produced. A 300 ml culture was then induced at an OD_{600} of 0.75 with 1 mM IPTG and were left to grow for 3~4 hr until OD_{600} ~1.6. Cells were then harvested at 6000 *g* at 4°C for 10 min and resuspended in 20 mM Tris-HCl (pH 8), 1 mM EDTA, 1 mM EGTA, 100 mM NaCl, 10% glycerol (v/v), 1 mM DTT, 1 mM PMSF, 1 tablet/100 ml of protease inhibitor cocktail (Roche). Cell pellets were sometimes frozen before resuspension at -20°C until ready for use. After resuspension, cells were incubated on ice for 30 min after addition of 1 mg/ml lysozyme. Sonication of the cells on ice at 10 μm for 5 min and extraction with 0.1% Triton X-100 (v/v) in dH_20 ensured most of the fusion protein was released from the cells. After 20 min of centrifugation, 12,000 *g*, 4°C, 15 ml of supernatant containing the expressed fusion protein was removed and added to 2 ml of 50% slurry of glutathione-agarose (Sigma) in PBS, 150 mM NaCl. This was incubated at 4°C for 1 hr on a roller mixer before loading into a column (Econo, Bio-

Rad). The supernatant was allowed to either drip through or be pumped through using a peristaltic pump (Gilson). The column was washed with 10 ml PBS-T (PBS, 1% Triton X-100 (v/v) and protease inhibitors as before (but no DTT). The GST fusion protein was eluted with 5 ml elution buffer (10 mM reduced glutathione [3 mg/ml] in 50 mM Tris-HCl [pH 8]) and protease inhibitors as before. Eluate was collected in 10x 0.5 ml fractions. A sample of each aliquot was mixed with an equal volume of 2x SDS loading buffer for protein analysis by SDS-PAGE (section 2.4.4) and visualised by coomassie staining (section 2.4.5). The yield of fusion protein was quantified against known BSA standards (10 mg/ml, NEB). Fractions containing the GST fusion protein were pooled and concentrated in centricon-10 columns (Amicon/Millipore) following manufacturer's instructions.

2.5.2 Immunisation

Host animals were tested for antibody manufacture suitability by western blotting (section 2.4.4) and immunofluorescence of their pre-immune serum on NIH 3T3 and HeLa cells (section 2.8.1). An animal from each species was subsequently chosen for immunisation with antigen, based on minimal cross-reacting proteins detected by the pre-immune on a western blot and low background fluorescence by immunofluorescence.

After ethical approval of the immunisation protocol by the Home Office, antigen was sent to Diagnostic Scotland (unconjugated peptide and 1 mg/ml of GST-fusion protein) for immunisation of the designated animals. Designated animals were firstly injected with a mix of synthetic peptide (0.5 mg/ml) and fusion protein (0.125 mg/ml). A subsequent booster four weeks later and two further booster injections were given and bleeds collected two weeks after each. A total of 2 mg/ml of peptide and 0.5 mg/ml of fusion protein were injected into each animal. The immune responses were checked both by western blotting (section 2.4.6) and immunofluorescence (section 2.8.1) with immunised serum on NIH 3T3 and HeLa cells.

2.5.3 Purification of antibodies

2.5.3.1 Immunoglobulin G purification of peptide and GST-fusion protein

Serum from bleed 3 was diluted 1:10 in 20 mM phosphate buffer, pH 7. 20 mM phosphate buffer was also passed through an Immunoglobulin G (IgG) HiTrap column (HiTrap™ Protein G HP, Amersham Biosciences), linked up to a peristaltic pump (Gilson) with the flow rate at 1 ml/min. 20 ml of diluted antibody was passed through the column and the column washed with 10 volumes of phosphate buffer to remove non-specifically bound proteins. Antibody was eluted with 3 ml of 100 mM glycine (pH 2.7) and 0.5 ml fractions collected into 1.5 M Tris-HCl (pH 8.8) to neutralise the eluate. 10 µl of each sample was mixed with 10 µl of 2x SDS loading buffer, boiled for 5 min and analysed by SDS-PAGE (section 2.4.4). 5 µl of each fraction was diluted in 500 µl PBS and the absorbance at A_{280} measured to estimate protein concentration given that an A_{280} of 1 equals 0.75 mg/ml of pure IgG. The peak fractions were pooled and antibody was stored at 4°C or -20°C in a final volume of 30% glycerol (v/v) with 1 mg/ml of BSA if the antibody concentration was less than 1 mg/ml.

2.5.3.2 Preparation of affinity columns for purification of GST-fusion antibodies

2.5.3.2.1 Binding of histidine-tagged fusion proteins on a nickel-agarose column

DNA sequence of the antigen was ligated into expression vector pET-32-a (Novagen), which produces a TRX (thioredoxin) fusion protein with a histidine (his)-tag to enable purification on a Ni-agarose column (HIS-Select™ Nickel Affinity Gel, Sigma). The plasmid was transformed into BL21-CodonPlus®(DE3)RP. A 300 ml culture at an OD_{600} of 0.75 was induced with 1 mM IPTG and incubated at 37°C for 3~5 hr ($0D_{600}$ ~1.6). Cells were pelleted at 6000 *g* at 4°C for 10 min and resuspended in 15 ml of binding buffer (0.3 M NaCl, 5 mM imidazole, 50 mm

NaH_2PO_4, 10% glycerol [v/v] and freshly added 10 mM β-ME and 0.1 mM PMSF). Cells were lysed by incubation with 1 mg/ml lysozyme for 30 min on ice, followed by sonication at 10 μm for 5 minutes on ice. Triton X-100 was added to 0.1% (v/v). After centrifugation at 20,000 *g* for 20 min at 4°C to remove bacterial debris, the supernatant containing the soluble fusion protein was removed for purification on a Ni-agarose column.

2 ml of 50% slurry of Ni-agarose in PBS was equilibrated by rotation with 10 ml of binding buffer for a minimum of 2 hr at 4°C. Centrifugation at 400 *g* for 5 min released the supernatant, which was removed and replaced with 15 ml of bacterial cell supernatant containing the his-tagged fusion protein. This mixture was rotated at 4°C for a minimum of 24 hr before being transferred into a small Bio-Rad column where the supernatant was allowed to drip through. The column was washed twice with 20 mM imidazole wash solution (20 mM imidazole, 0.3 M NaCl, 50 mM NaH_2PO_4, 0.1% Triton X-100 [v/v], 10% glycerol [v/v], 10 mM β-ME, 0.1 mM PMSF) and the fusion protein eluted in 1 ml fractions with 5 ml of 250 mM imidazole buffer (250 mM imidazole, 0.3 M NaCl, 50 mM NaH_2PO_4, 0.1% Triton X-100 [v/v], 10% glycerol [v/v], 10 mM β-ME and 0.1 mM PMSF). Samples were analysed as in 2.5.1 and concentrated in a centricon-10 column.

2.5.3.2.2 Desalting and buffer exchange of his-tagged fusion protein on a PD-10 column

Equilibration buffer supplied with the PD-10 column (Amersham Biosciences) was allowed to run into the column by removal of the end caps. His-tagged fusion protein was added to the column bed in a sample volume made up to 2.5 ml with coupling buffer (0.1 M NaHCO3, 0.5 M NaCl at pH 8.3). When the sample had run into the bed, high molecular weight proteins (>5 kDa) were separated from low molecular weight substances (<1 kDa) by molecular exclusion chromatography and eluted off the column in 7x 0.5 ml fractions with 3.5 ml of coupling buffer. The exchange into coupling buffer allows fusion protein to be coupled to CNBr-activated Sepharose™ 4B via its amino groups, which are now predominantly in the unprotonated (desalted) form (section 2.5.3.2.3). 10 μl of each fraction was

analysed by boiling for 5 min with 10 µl of 2x SDS loading buffer and analysed by SDS-PAGE (section 2.4.4).

2.5.3.2.3 Making a CNBr-activated Sepharose™ 4B affinity column

The column was prepared by resuspending 1 *g* of CNBr-activated Sepharose™ 4B (Amersham Biosciences) in 15 ml of 1 mM HCl. for 15 min on a mixing rotor. 200 ml of 1 mM HCl/g of gel matrix was passed through the sepharose 4B matrix whilst linked up to the peristaltic pump and constantly mixed to allow complete washing of the matrix. The matrix was centrifuged in 1 mM HCl at 12,000 *g* and its vol measured in a 15 ml falcon tube. 50% (w/v) slurry of matrix and 1 mM HCl was stored at 4°C until required.

To make a 1 ml column, 2 ml of slurry was resuspended in a 15 ml falcon with 10 ml coupling buffer, mixed by gentle inversion and centrifuged at 400 *g* for 1 min. The coupling buffer was aspirated and the matrix resuspended with approximately 5 mg (1.5 ml) of desalted fusion protein from the PD-10 column and 1.5 ml coupling buffer (total volume added to matrix was 3 ml). The column was mixed by gentle inversion at room temperature for 1 hr per 1 ml of gel. After centrifugation, the matrix was washed for a further 5 min in 14 ml of coupling buffer. A 2 hr wash at RT with 14 ml of 100 mM Tris-HCl (pH 8) deactivates the remaining active groups on the column. Subsequent alternate washing with 0.1 M NaCOOH3 (pH 4) containing 0.5 M NaCl, followed by 0.1 M Tris-HCl (pH 8) containing 0.5 M NaCl, ensures removal of excess fusion protein so that no protein remains ionically bound to the immobilised 492 fusion protein. The matrix was poured into a Bio-Rad disposable column and held upright with support to produce the affinity column. The column was stored in storage buffer (0.05% NaN_3 in PBS [w/v]) until required.

2.5.3.3 Affinity purification of GST-fusion antibody

For a 1 ml column, 2.5 ml of serum was diluted up to 25 ml in PBS (1:10 dilution) and filtered through a 0.45 µm filter. 10 bed volumes (i.e. 10 ml) of PBS were run through the affinity column before loading the column with diluted serum. Flow through was collected to check for non-specific binding of proteins. Another 10 ml

of PBS was used to wash the column followed by 20 bed volumes of High Salt wash buffer (10 mM Tris-HCl [pH7.5], 500 mM NaCl). The column was washed with 20-50 bed volumes of PBS until ready to elute antibody. 35 µl of 1 M Tris-HCl (pH 8.8) was measured into 12-15 1 ml collection tubes and upon addition of elution buffer (100 mM glycine [pH 2.7], 150 mM NaCl, 10% glycerol [w/v]), 0.5 ml fractions were collected, mixed and placed on ice. 5 µl of each fraction was diluted in 500 µl PBS and the absorbance at A_{280} measured to estimate protein concentration (section 2.5.3.1). The peak fractions were pooled after analysis on an acrylamide gel. Antibody was stored at 4°C or -20°C in a final volume of 30% glycerol [w/v] with 1 mg/ml of BSA if the antibody concentration is less than 1 mg/ml.

2.5.3.4. Preparation of affinity column for purification of peptide antibody

2.5.3.4.1 Reduction of unconjugated peptide

A Reduce-Imm™ Immobilised Reductant Column and Reducing Kit (Pierce) were used according to manufacturer's instructions to reduce any disulphide bonds present in the unconjugated peptide (KLH free). This ensured that any sulfhydryl (–SH) groups occurring in the side chain of cysteine (cys) residues were free for conjugation and immobilisation of the peptide on a gel support. A Reduce-Imm™ Immobilized Reductant Column was equilibrated at RT with 5 ml of the supplied equilibration buffer. The column was activated by addition of 10 ml of 10 mM DTT. Non-immobilised DTT was removed from the column after washing with 10 ml of equilibration buffer. 2.4 ml (10 mg) of unconjugated peptide was added to the column and incubated for 1 hr. The reduced peptide was recovered from the column by applying 9 ml of equilibration buffer and collecting 1 ml fractions. Fractions containing the reduced peptide were determined using 10 mM Ellman's Reagent supplied with the kit. Briefly, 20 µl of each fraction was added to 1 ml of equilibration buffer and 100 µl of Ellman's Reagent added upon which, fractions containing reduced peptide turned yellow. After incubation for 15 min, an A_{412} reading was taken and fractions containing the peak absorbance were pooled.

The Reduce-Imm™ Immobilized Reductant Column was regenerated by repeating the activation step with 10 mM DTT and subsequently stored at 4°C after a 10 ml wash in PBS containing 0.05% NaN_3 (w/v).

2.5.3.4.2. Immobilisation of reduced peptide on SulfoLink® Coupling Gel

5 ml of SulfoLink® Coupling Gel slurry was placed into a disposable Bio-Rad column and equilibrated at RT with 4 column volumes of coupling buffer (50 mM Tris-HCl, 5 mM EDTA-Na [pH 8.5]). 3 ml of sulfhydryl-containing peptide (2.5.3.4.1) diluted with 2 ml of coupling buffer (total volume = 5 ml) were added to the column and the column mixed at RT for 15 min. The column was then incubated for 30 min without mixing whilst the gel bed settled. The column was washed with 3 column volumes of coupling buffer after which the non-specific binding sites on the column were blocked by addition of 5 ml of 50 mM L-cysteine•HCl in coupling buffer with gentle mixing for 15 min at RT followed by an incubation for 30 min without mixing. 6 column volumes of 1 M NaCl was used to wash the column followed by washing with 2 column volumes of storage buffer (0.05% NaN_3 in PBS [w/v]) before storage of the affinity column in a further 2 ml of storage buffer.

2.5.3.5 Affinity purification of peptide antibody

For a 5 ml column, 10 ml of serum were diluted up to 50 ml in PBS (1:5 dilution) and filtered through a 0.45 µm filter. 6 ml of PBS were run through the affinity column (2.5.3.4.2) before loading the column with diluted serum. The serum was passed over the column 4 times before washing with 3 bed volumes of PBS. 100 µl of 1M Tris-HCl (pH 7.5) was measured into 8 x 1.5 ml collection tubes and upon addition of 8 ml of elution buffer (100 mM glycine [pH 2.7]), 1 ml fractions were collected, mixed and placed on ice. 10 µl of each fraction was mixed with 10 µl of 2x SDS loading buffer, boiled for 5 min and analysed by SDS-PAGE (section 2.4.4). The peak fractions were pooled after analysis on an acrylamide gel. Antibody was stored at 4°C or -20°C in a final volume of 30% glycerol (w/v) with 1 mg/ml of BSA if the antibody concentration is less than 1 mg/ml.

2.6 Transgenic animal production

2.6.1 Animal husbandry

All animals were maintained in a specific pathogen free (SPF) environment and experiments were carried out under home office licence. Wild-type (wt) mice (C57BL6 [Charles River, UK] or MFI [Harlan]) were bred in-house with mice heterozygous for the transgene to establishment gene-trap knockout lines. Embryos for all experiments were generated from timed matings with females from the outbred mouse line CD1 (Charles River, UK) or transgenic females with the morning of vaginal plug detection being counted as embryonic day 0.5 (E0.5).

2.6.2 Blastocyst injection of embryonic stem cells and establishment of transgenic mouse lines

The gene-trap screen by Sutherland *et al.* in 2001 was performed with passage 19 (pg19) E14 embryonic stem (ES) cells, which are derived from the 129/Ola mouse strain. Between 10-50 cells from cell lines ES113 at pg6 and ES492 at pg5 were injected into 2 different female C57BL6 blastocyst recipients at 2.5 days post coitum (dpc) by Shelia Webb of the MRC Human Genetics Unit in Edinburgh and these blastocysts were placed into pseudopregnant C57BL6 females by uterine transfer. Chimeras among the offspring were identified by a mottled coat colour and these were bred with both outbred MF1 or inbred C57BL6 mice to determine if the mutant ES cell lines were transmitting through the germ line to establish mutant lines. Resulting progeny were tested for the presence of the pGT1-3 transgene by PCR amplification using primers that amplify a 413 bp (base pairs) fragment within the *LacZ* gene (section 2.2.12). Heterozygous lines for both mice were established by repeatedly backcrossing *LacZ* positive mice with either wt MF1 or C57BL6 mice. *LacZ* positive mice were bred for congenic inbred and outbred lines.

2.6.3 Harvesting of post-implantation embryos

Post-implantation embryos were harvested for wholemount X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) (Melford) staining and/or paraffin wax sectioning. The plugged females were sacrificed by cervical dislocation at the appropriate time point and embryos removed from their uteri into PBS. Embryos were freed from the extra-embryonic membranes and yolk sac using scissors and forceps, and rinsed in fresh PBS prior to subsequent fixation (section 2.8.4.1).

2.7 Mammalian Cell Culture

2.7.1 General methodology

2.7.1.1 Freezing and thawing cells stored in liquid nitrogen

All cell suspensions were frozen in 0.5 ml aliquots of 7% DMSO (v/v) in foetal calf serum (FCS) (except mouse ES cells which were frozen in 10% DMSO/FCS) before being stored in cryotubes in frozen liquid N_2. To retrieve cell lines from liquid N_2, cells were thawed at 37°C for 5 min in a water bath and were spun down in culture medium to remove the DMSO before seeding into T25 cm^2 culture flasks. Mouse ES and F9 embryonic carcinoma (EC) cells were seeded into T25 cm^2 culture flasks coated with 0.1% gelatin (Sigma, G-1890), prepared by firstly layering a solution of 0.1% gelatin in PBS over the bottom of the flask and aspirating off before adding cells and media.

2.7.1.2 Routine cell culture and harvesting

All cells were incubated at 37°C with 5% CO_2 and tissue culture medium as described in Table 2.3. Cells other than mouse ES cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen™) and 10% FCS (v/v) containing penicillin (1000 U/ml) and streptomycin (650 μg/ml). ES cells were grown in Glasgow Modified Eagle Medium (GMEM, Invitrogen™) with supplements (Table 2.6) as described by Smith *et al.* in 1988. Cells were grown to near confluence before

splitting into fresh tissue culture flasks coated in gelatin if required. To split the cells, medium was poured off, flasks were rinsed with PBS and monolayers were then covered with 1.5 ml of trypsin-EDTA and incubated at 37°C for 5 min. Gentle agitation dislodged the cells, fresh medium was added, and the cells pelleted at 1000 *g* for 5 min before being replated or harvested for experiments.

Cells were counted (section 2.7.1.3) before being seeded at the relevant density onto microscope slides for immunofluorescence (ES cells, 10^6 cells per slide). Slides were prepared by soaking in 100% ethanol before being left to dry in QuadriPerm slide chambers (VIVASCIENCE, Satorius) inside a laminar flow sterile fume hood (Heraeus). The slides were coated with gelatin for growth of ES or EC cells.

2.7.1.3 Cell counting

For cell counting, cells were resuspended in PBS after harvesting and a drop placed on a haemocytometer (Weber Scientific International Ltd.) with a weighted coverslip sealed on top. The total volume defined by the grid was $1x10^{-4}$ ml and cell concentrations per ml were obtained by multiplying the total number of cells over the grid by 10^4.

Table 2.6 General information regarding mammalian cell lines

Name	*Description*	*Sub-culture*	*Medium*	*Source*
ES cells:				
E14	wt mouse ES cell, (129/Ola). In house.	1:5 every other day	GMEM, 10% FCS (v/v), 100 U/ml LIF (Leukaemia inhibitory factor, pers. com.), 2 mM glutamine, 50 μM β-ME, 100 mM sodium pyruvate (Sigma), 1% NEAA (non-essential amino acids) (v/v) (Sigma)	gift, A. Smith (Hooper *et al.*, 1987)
Suv39h wt	wt ES cells (C57BL6/129)	"	"	gift, T. Jenuwein (Peters *et al.*, 2001)
Suv39h dn	*Suv39h1* and *Suv39h2* null function mouse ES cells	"	"	"
Dnmt3ab wt	J1 wt ES cells (129S4/SvJae)	"	"	gift, E. Li (Okano *et al.*, 1999)
Dnmt3ab$^{-/-}$	Targeted disruption of *Dnmt3a* and *3b* in mouse J1 ES cells	"	"	"

OS25 wt	*Oct4-HygTK*, *Sox2βgeo* into E14Tg2a. 40XY. (129/Ola)	"	"	gift, A. Smith (Billon *et al.*, 2002)
Other cells:				
NIH 3T3	Mouse fibroblast	1:10 every 4-6 days	DMEM, 10% FCS (v/v)	In house
HeLa	Human, epithelial carcinoma	1:3 – 1:10	"	In house
F9	Mouse EC (male testicular derived teratocarcinoma)	1:10-1:20 every 2-4 days	"	In house
M15	Mouse mesonephric	1:5 - 1:10	"	gift, Anna Shafe
LA9	Mouse fibroblast	1:5	"	gift, Ming Hong
C127	Mouse mammary gland tumour	1:5	"	In house
P19	Mouse EC derived from teratocarcinoma	1:3 to 1:6	DMEM, 7.5% FCS (v/v), 1% NEAA (v/v)	In house
ATDC5	Mouse EC derived from chondrogenic teratocarcinoma cells	1:8 every 2-3 days	DMEM:Ham's F12 (1:1) (Sigma), 2 mM glutamine, 5% FCS (v/v), 10 μg/ml human transferrin (Sigma), $3x10^{-8}$ M sodium selenite (Sigma)	ECACC (European Collection of Cell cultures)

2.7.2 *In vitro* differentiation of ES cells (Strickland and Mahdavi, 1978) and F9 cells

Cells were passaged as described (section 2.7.1.2), seeded (10^4 cells/cm^2) onto gelatin-coated flasks and were grown overnight in normal media. The next day, cells were rinsed with PBS and the media changed to LIF free media (ES cells) or normal media (F9 cells) supplemented with 5x 10^{-6} M retinoic acid (all-*trans*-retinoic acid, stock solution 3.3x 10^{-3} M in DMSO, stored in the dark at -20°C). The retinoic acid, containing medium was renewed every 24 hr for a period of two (F9 cells) or four days (ES cells) following induction.

2.7.3 *In vitro* differentiation of ATDC5 cells

To culture undifferentiated ATDC5 cells, cells were seeded at 6x10^4 cells per 6 well plate in maintenance media (Table 2.6). Cells proliferate rapidly until confluent (1.4x10^5 cells/cm^2) where they remain undifferentiated with a fibroblastic morphology (Shukunami *et al.*, 1996). To differentiate ATDC5 cells, cells were seeded at the same density per 6 well plate but this time cultured in maintenance medium supplemented with 10 mg/ml bovine insulin (Shukunami *et al.*, 1996). By day 3 of differentiation cells stop growing by contact inhibition and by day 6-10, cells have condensed into an elongated spindle-like morphology with a decreased growth rate. Day 14-21 sees the formation of nodules by proliferating cells with a round morphology. Cells around the nodules remain fibroblastic and by day 24 cartilage nodules have stopped growing.

2.7.4 *In vitro* differentiation of OS25 cells (Billon *et al.*, 2002)

Undifferentiated OS25 cells were cultured in normal LIF containing ES cell culture medium in the presence of 100 μM G418 to select for undifferentiated cells. To differentiate OS25 cells, the differentiation protocol as described by Billon *et al.*, 2002, was modified as described by Chambeyron and Bickmore, 2004. Cells were grown for 1 day in a T25 cm^2 flask without LIF and then grown for 2 days with

retinoic acid (5x 10^{-6} M) present in the LIF free media. Cells were then harvested in 4 ml of media before being seeded onto microscope slides at 1 ml of cell suspension per slide. Cells were left to grow on the slides for two days with gancyclovir (2.5 μM) to select against undifferentiated cells. The cells were left to grow on the slides for a further four days in gancyclovir alone with the media being renewed every other day.

2.7.5 Transfections

2.7.5.1 NIH 3T3 cells

$2.4x10^5$ cells were seeded onto coverslips in 6 cm^2 dishes and grown to 70% confluence. For each transfection, 5 μl Lipofectamine 2000 (Invitrogen™) was added to 200 μl of Opti-MEM (Invitrogen™), vortexed and incubated for 5 min. A further 200 μl of Opti-MEM was incubated with 2 μg DNA for 5 min. Both incubation mixtures were added together and left to incubate for a further 20 min. Media was removed from the cells and 3 ml of fresh media added with drop-wise addition of the transfection mix. Cells were incubated for 24 hr and then fixed for immunofluorescence (section 2.8.1) or visualised directly under a fluorescent microscope for GFP (green fluorescent protein) expression.

2.7.5.2 ES cells

ES cells were seeded in the morning onto gelatinised coverslips at $3.6x10^5$ cells per 6 cm^2 dish ready to transfect that afternoon or alternatively they were seeded at $7.8x10^4$ cells per dish ready to transfect the following day. 2 μg of DNA was added to 76 μl Opti-MEM and mixed for 5 min during which time 2 μl of Lipofectamine2000 (Invitrogen™) was mixed with 76 μl Opti-MEM and incubated for 5 min. The above solutions were mixed and left to incubate for 20 min before being adding drop-wise to the cells. Cells were left to transfect overnight and analysed the following morning by fluorescent microscopy.

2.8 Immunohistochemistry

2.8.1 Immunofluorescence on fixed cells

Mammalian cells were grown on slides as described (section 2.7.1.2) and all subsequent incubations performed at RT. Slides were rinsed in PBS containing 1.5 mM $MgCl_2$ and 1 mM $CaCl_2$ and fixed for either 10 or 20 min in 4% or 3% paraformaldehyde (pFa) (w/v) in PBS respectively. All subsequent steps were performed with PBS containing Mg and Ca. After a further 3 washes in PBS, the slides were quenched in 50 mM NH_4Cl in PBS for 10 min and permeabilised for 12 min in 0.1% Triton X-100 (v/v) in PBS. Slides were washed a further 3 times in PBS before blocking in 5% donkey serum (v/v) in PBS for 20 min. Slides were then incubated o/n in moistened chambers with primary antibody (diluted in 5% block) under a parafilm coverslip (Table 2.7). After washing, slides were incubated in secondary antibody (Table 2.8) diluted in the same way but this time for 1 hr. All secondary antibodies were species-specific fluorescein isothiocyanate (FITC) or Texas Red (TR) conjugates obtained from the Jackson Laboratories or Vector Lab. Secondary antibody was washed as above and all slides were mounted with 0.5 μg/ml 4,6-diamidino-2-phenylindole (DAPI) in Vectashield (Vector). Coverslips were sealed with rubber solution (PANG) and slides were stored in the dark at 4°C until imaged (section 2.10.3). For dual epitope IF with primary antibodies that had been raised in the same species (i.e. rabbit), the Zenon Rabbit IgG labelling kit (Molecular Probes) was used to label one of the two antibodies according to the manufacturer's instructions. After primary antibody incubation with the first antibody was performed as for normal IF, the secondary antibody incubation then followed. Instead of proceeding to mount the slides, 1 μg of the second primary antibody was labelled with 5 μl of Zenon rabbit labelling reagent (FITC labelled Fab fragments) and the volume made up to 10 μl with PBS. After incubation for 5 min, 5 μl of blocking reagent (non-specific IgG) was subsequently added to bind any remaining unbound Fab fragments. The total reaction volume was not allowed to exceed 20 μl and the labelled antibody was applied to the slide within 30 min, after removing the secondary antibody by washing in PBS as before. A second fixation

step of 10 min in 4% pFa (w/v) followed, and after a PBS wash, slides were mounted as for normal IF.

Table 2.7 Primary antibody used for immunofluorescence

Antibody	*Species*	*Source*	*Dilution Factor*
Anti-KAP-1	Mouse monoclonal	David Schultz	neat
Anti-β-gal	Rabbit polyclonal	Europa	1:2000
Anti-NT-2	Rabbit polyclonal	Yoshihiko Yamada	1:200
Anti-492Ab IgG	Rabbit polyclonal	This thesis	1:1000
Anti-492Ab AP	Rabbit polyclonal	"	1:100
Anti-BrdU	Rat monoclonal	Harlan Seralab	1:100
Anti-SSEA-1	Mouse polyclonal	DSHB	1:200
Anti-fibrillarin/ Nop1p (mAb D77)	Mouse monoclonal	gift, J. Aris (Florida, USA)	1:1000
Anti-HP1α	Mouse monoclonal	Chemicon Int.	1:500
Anti-HP1β	Mouse monoclonal	Chemicon Int.	1:500
Anti-HP1γ	Mouse monoclonal	Chemicon Int.	1:500
CREST (Campbell)	Human anti-sera	B. Sullivan	1:300
Anti-2b-Met_3H3-K9 (rabbit no. 4861)	Rabbit polyclonal	Peters *et al.*, 2001	1:500
PML	Mouse monoclonal	Chemicon	1:100

Table 2.8 Secondary antibodies used for immunofluorescence

Antibody	*Species*	*Source*	*Dilution Factor*
Anti-rabbit FITC conjugate (IgG, heavy & light chain specific [H&L])	Donkey	Jackson Labs, 711-095-152	1:150
Anti-rabbit TR conjugate (IgG, H&L)	Donkey	Vector Labs	1:100
Anti-mouse FITC conjugate (IgG, H&L)	Donkey	Vector Labs	1:100
Anti-mouse TR conjugate (IgG, H&L)	Donkey	Jackson Labs, 715-075-150	1:100
Anti-rat TR conjugate (IgG, H&L)	Donkey	Jackson Labs, 711-075-152	1:100
Anti-sheep FITC conjugate (IgG F[ab']2 [H+L])	Donkey	Jackson Labs, 313-096-047	1:75

2.8.2 Immunofluorescence on fixed cells with bromodeoxyuridine incorporation

For immunofluorescence of bromodeoxyuridine (BrdU) incorporation into DNA synthesising cells, BrdU was added 30 min before 3% pFa (w/v) fixation at a concentration of 0.01 M (Boehringer). Cells were fixed in 10% formalin (v/v) for 10 min and lysed with 0.1% Triton X-100 (w/v) for 12 min. A 30 min 2 M HCl treatment denatures the DNA in the cells for antibody accessibility. Cells were washed in PBS, blocked for 10 min in 5% BSA (w/v) before incubating in anti-Brd-

U antibody (Table 2.7) for 1hr. Secondary antibody incubation and mounting of the slides in DAPI/Vectashield was performed as described above.

2.8.3 Detection of the differentiation state of cells

2.8.3.1 Alkaline phosphatase staining of ES cells

The differentiation state of ES cells were detected using an alkaline phosphatase assay performed using a kit (Sigma, 86-R) containing a fixative and stain. Culture medium from the cells grown in 6-well plates was replaced with 1.5 ml fix and left for 30 sec. The fix was removed and the cells rinsed gently with dH_20 for 45 sec. Cells were rinsed again and enough stain (prepared fresh) was gently poured into each well. Cells were incubated in the dark for 15 min at RT. Undifferentiated ES cell colonies stain intensely pink. Stain was removed and the wells rinsed with water and air-dried.

2.8.3.2 Alcian blue staining of ATDC5 cells

Mouse ATDC5 cells were grown on slides or coverslips as described (section 2.5.5) and rinsed with PBS. Cells were fixed for 20 min in 95% methanol (v/v in dH_20) and stained o/n with 0.1% alcian blue 8GS (w/v) in 0.1M HCl (pH 1). Stain was rinsed off in PBS and the slides photographed (section 2.10.2).

2.8.4 Wholemount expression analysis of mouse embryos and tissues

2.8.4.1 Tissue fixation

1 M phosphate buffer (pH 7.3) was prepared by mixing 1M NaH_2PO_4 and 1 M Na_2HPO_4, in a ratio of 21:6. Mouse embryos or tissues were fixed for 1 hr in either 10 ml of gluteraldehyde fixative (5 mM EGTA [pH 8], 0.2% gluteraldehyde [v/v] and 2 mM $MgCl_2$ in 0.1 M phosphate buffer) or 4% pFa (w/v) immediately after dissection (section 2.4.3). Embryos were washed twice at RT for 15 min in wash

solution (0.05% BSA [w/v], 2 mM $MgCl_2$, 0.02% NP-40 (v/v), 0.1% sodium desoxycholate [w/v] in 0.1 M phosphate buffer) before staining.

2.8.4.2 X-Gal staining

X-Gal staining solution was prepared by fully dissolving 250 mg X-Gal in 5 ml dimethylformamide (DMF) before adding to the rest of the X-Gal reaction buffer (5 mM potassium ferricyanide [$K_3Fe(CN)_6$], 5 mM potassium ferrocyanide [$K_4Fe(CN)_6$], 0.25 mg/ml spermidine, 2 mM $MgCl_2$ and made up to 500 ml in wash solution [2.8.4.1]). This solution was filtered before use. 50 ml aliquots were stored at -20°C for later use. After fixation and washing, embryos were stained o/n in X-Gal staining solution (maximum 15 hr) at RT whilst protected from the light. Staining solution was removed the following morning and the samples stored in fix until required. Embryos were visualised with microscopy (section 2.10.1).

2.8.5 Mouse histology

2.8.5.1 Tissue processing for paraffin wax sectioning

For histochemistry, embryos were processed for paraffin wax embedding and microtome sectioned. Tissue processing was automated through 11 stages of equal duration (usually 30-40 min depending on the sample size) using a Tissue-Tek™ VIP, at the settings presented in table 2.9.

Table 2.9 Settings used for wax processing of embryos/tissues

Stage	*Temperature (°C)*
70% ethanol (v/v) in PBS	40
85% ethanol/PBS	40
95% ethanol/PBS	40
100% ethanol/PBS	40
100% ethanol/PBS	40
xylene	60
xylene	60
paraffin wax	60
paraffin wax	60
paraffin wax	60
paraffin wax	60

For fragile tissues, the pressure vacuum was switched on at the second 100% ethanol stage to maintain the integrity of the tissue during the embedding stages. Smaller embryos less than 9 dpc, were hand embedded in wax to minimise damage to the embryos. A procedure similar to the above was followed but the duration of each stage was 15 min. Following the final wax stage, samples were transferred into a suitable mould, orientated and allowed to solidify in the wax bed. The wax block was trimmed to the desired size and the edges angled to improve sectioning on the microtome.

2.8.5.2 Microtome sectioning, dewaxing and tissue staining

Wax embedded embryos were sectioned at 7-10 μm on a microtome and ribbons of sections floated out in a 42°C water bath. Sections were transferred to Superfrost Plus slides (BDH) and the slides incubated overnight at 55°C to allow sections to dry. Slides were dewaxed by washing for 3x 10 min in fresh xylene, followed by 3x 10 min washes in 100% ethanol. Slides were then passed through a graded ethanol series, washed in dH_20 and counterstained with nuclear fast red (nfr) according to a

standard protocol. Typically, slides were immersed in nfr for 1 min before being brought back up the alcohol series into 3 final washes in Histoclear™ (national diagnostics, US). Sections were mounted in Histomount™ (national diagnostics, US) to preserve X-Gal staining.

2.9 Fluorescence *in situ* hybridisation

2.9.1 Fixing of cells in 3:1 methanol: acetic acid

Cells for two-dimensional (2-D) fluorescence *in situ* hybridisation (FISH) were prepared as follows: cells were harvested and resuspended in 10 ml hypotonic solution (0.033 M KCl, 0.017 M tri-sodium citrate), which was added drop-wise with constant agitation (the concentration of cells in hypotonic should be $<2x\ 10^7$/ml). The cells were left to swell for 10 min at RT before centrifugation at 400 *g* for 5 min. Cells were then fixed in fresh 3:1 methanol:glacial acetic acid (MAA), again added drop-wise with constant agitation before incubating on ice for 20 min. After centrifugation, the cells were resuspended in 8 ml of fix and the cells placed at 4°C o/n. Cells were fixed twice more and stored indefinitely at -20°C.

2.9.2 Preparation of probes for FISH

2.9.2.1 Nick translation

DNA was labelled using biotin-16-dUTP or digoxigenin-11-dUTP incorporation by nick translation. 1-1.5 µg of DNA was added to 4 µl 10x nick translation salts (0.5 M Tris-HCl [pH7.5], 0.1 M $MgS0_4$, 1 mM DTT, 500 µg/ml BSA), 4 µl each of 2 mM dATP, dGTP and dCTP, 2 µl of 0.5 mM dTTP and 4 µl biotin-16-dUTP or digoxigenin-11-dUTP (Roche). DNase I was added to a final concentration of 1 U/ml along with 1 µl of T4 DNA polymerase I (Invitrogen™, 10 U/µl). The total volume was made up to 40 µl with dH_20 and mixed thoroughly before being incubated at 16°C for 90 min. The reaction was terminated by placing at -20°C or immediately processed for the removal of unincorporated label (section 2.9.2.2).

2.9.2.2 Removal of unincorporated label

Quick Spin columns (Roche) containing G-50 Sephadex beads were used in accordance with the manufacturer's instructions to remove any free biotin-16-dUTP, digoxigenin-11-dUTP or dNTPs remaining in the solution. Cleaned probes were eluted in 40 µl TE (pH 8).

2.9.2.3 Quantification of label incorporation

Gridded nitrocellulose membranes were prepared by brief soaking in dH_20 followed by 20x SSC for 10 min. Labelled DNA probes were diluted to $1x10^{-3}$ and $1x10^{-4}$ in TE (pH 8) and 1 and 2 µl of each were spotted onto the gridded membrane. On the same membrane 20, 10, 2 and 1 pg of labelled lambda DNA standards (Roche) were also spotted. DNA was cross-linked onto the membrane by exposure to 30 mJ of UV irradiation.

The membrane was immersed in buffer 1 (0.1 M Tris-HCl [pH7.5], 0.15 M NaCl) for 5 min at RT, then blocked in 5% Marvel (w/v) in buffer 1 at 37°C for 30 min. 10 µl streptavidin-alkaline phosphatase (Boehringer) and/or anti-digoxigenin-alkaline phosphatase (Boehringer) were added to 10 ml of buffer 1 and placed in a sealed polythene bag with the membrane for 30 min at RT. The membrane was washed twice in buffer 1 then equilibrated for 5 min in 0.1 M Tris-HCl (pH 9.5). The colour reaction was developed by incubation of the membrane in a sealed polythene bag, with 5 ml of 0.1 M Tris-HCl (pH 9.5) and two drops from bottles 1-3 from the alkaline phosphatase substrate kit IV (Vector). The substrates in this colour reaction are 5-bromo-4-chloro-3-indolyl phosphate and nitroblue tetrazolium, which produce a blue reaction product. A complete colour reaction was observed within a few hours of incubation at RT in the dark and an estimate of the concentration of DNA labelled probe was made by comparison with known lambda standards.

2.9.3 The FISH protocol for MAA fixed nuclei

2.9.3.1 Slide preparation

Glass slides were stored in a dilute solution of HCl in ethanol and were dried and polished with muslin before use. MAA fixed cells (section 2.6.1) were removed from storage at -20°C and centrifuged at 400 *g* for 5 min. Fresh MAA fix was added until the cell suspension reached a 'milky' appearance. One drop of suspension from a fine tipped pastette was dropped onto a horizontal microscope slide from a 30 cm height (the best chromosomal spreads were achieved when the air humidity was ~50% and by breathing on the slides initially). The quality of the spread was monitored by phase contrast microscopy. Slides were stored for 2-6 days prior to hybridisation.

2.9.3.2 Hybridisation

Slides were treated with 100 µg/ml RNaseA in 2x SSC for 1 hr at 37°C, washed briefly in 2x SSC and dehydrated through an ethanol series (2 min each in 70%, 90% and 100% ethanol). Slides were dried under a vacuum for 10 min before being heated in a 70°C oven for 5 min and immediately denatured in 70% formamide (v/v) in 2x SSC (pH 7.8) at 70°C for 1-2 min. Slides were plunged in 70% ethanol at 4°C for 2 min before dehydration through 90% and 100% ethanol.

Labelled probes (section 2.9.2) were prepared by precipitation of ~75 ng probe (200 ng BAC probe) with 5 µg salmon sperm DNA and mouse Cot 1 DNA (Invitrogen™, 2.5-10 µg depending on repeat content of probe). After the addition of 2x vol of ethanol, probes were spun down under a vacuum before resuspension in 10 µl hybridisation mix (50% deionised formamide [v/v], 10% dextran sulphate [v/v], 1% Tween 20 [v/v], in 2x SSC) or 13 µl of commercial mouse chromosome paints supplied in hybridisation buffer (Cambio). All probes were denatured at 70°C for 5 min and reannealed at 37°C for 15 min before being spotted onto coverslips and picked up by the slides. Slides were sealed with rubber solution (TipTop) before incubation o/n in a covered tray in a 37°C water bath.

2.9.3.3 Washing and detection of FISH signal

Rubber solution was removed from the slides and they were immersed in 2x SSC at 45°C for 4x 3 min. The coverslips fall off naturally. Slides were washed a further 4x 3 min in 0.1x SSC at 60°C before transfer to 0.1% Tween 20 [v/v] in 4x SSC. Detection was carried out in a moist chamber pre-heated to 37°C. Biotin was detected with sequential layers of fluorochrome-conjugated avidin (FITC- or TR-avidin), biotinylated anti-avidin, and a further layer of fluorochrome-conjugated avidin. Digoxigenin was detected with sequential layers of Rhodamine (R)-conjugated anti-digoxigenin and TR-conjugated anti-sheep IgG. Detection reagents were diluted in SSCM (4x SSC, 5% Marvel milk powder [w/v]) to the appropriate concentration (Table 2.10). After blocking with 40 µl of SSCM for 5 min at RT, 40 µl of the appropriate detection layer was applied to each slide. Slides were incubated in the same way at 37°C for 60 min followed by 3x 2 min washes of 0.1% Tween 20 [v/v] in 4x SSC at 37°C. All slides were mounted in 0.5 µg/ml of DAPI in Vectashield. Coverslips were sealed with rubber solution (PANG) and slides were stored in the dark at 4°C until imaged.

Table 2.10 Antibodies and fluorochrome-conjugates used for FISH

Antibody or fluorochrome-conjugate (cell sorting grade)	*Species in which the antibody/ conjugate was raised*	*Source*	*Stock concentration (mg/ml)*	*Dilution*
FITC-avidin	Goat	Vector	2.0	1:500
TR-avidin	Goat	Vector	2.0	1:500
Biotinylated anti-avidin	Goat	Vector	0.5	1:100
R-anti-digoxigenin	Sheep	Roche	0.2	1:20
TR-anti-sheep (IgG, H&L)	Rabbit	Vector	0.5	1:100

2.10 Fluorescence and brightfield imaging and processing

2.10.1 Wholemount microscopy

Dissections of post implantation embryos were imaged with the aid of a Leica Stereo MZFLIII stereofluorescence microscope (Leica Microsystems, Milton Keynes, UK) fitted with a fibre optic cold light source and illuminated base for incident/transmission illumination. Images of whole embryos under PBS were captured with a Photometrics CoolSnap colour CCD camera (Roper Scientific, Arizona) controlled scripts written for IPLab Spectrum (Scanalytics Inc., VA).

2.10.2 Brightfield analysis of cells or tissue sections

Embryo sections and cells were analysed with brightfield (using differential interface contrast [DIC] optics) and images taken with a Photometrics CoolSnap HQ monochrome CCD camera (Roper Scientific, Arizona). IPLAB Spectrum (Scanalytics Inc., VA) wrote the image capture scripts that controlled camera capture.

2.10.3 Fluorescence imaging of cells or nuclei

2-D FISH slides or immunofluorescence experiments were examined in a methodical manner using either a Zeiss Axioplan II or Zeiss Axioplan fluorescence microscope both with 100 watt mercury bulbs and equipped with a triple band-pass filter (Chroma # 83000). Grey scale images for each fluorochrome were collected with a cooled CCD camera depending on the model of microscope (Pentamax with a Kodak KAF 1400 sensor or Micromax with Kodak KAF 1400e sensor respectively, Princeton Instruments) using IPLAB software v. 3.6 (Scanlytics, USA).

2.11 Computational methods

The Bioinformatics programmes and resources referred to in this thesis are listed below, together with the relevant World Wide Web (www) link.

Ensembl	http://www.ensembl.org/
ExPASy-Tools	http://ca.expasy.org/tools/
InterProScan	http://www.ebi.ac.uk/InterProScan/
InterPro (proteome analysis)	http://www.ebi.ac.uk/integr8/ProteomeAnalysisAction
NCBI (BLASTN/P, TBLASTN)	http://www.ncbi.nih.gov/BLAST/
NCBI (Entrez)	http://www.ncbi.nlm.nih.gov/Entrez/index.html
NPD (nuclear protein database)	http://npd.hgu.mrc.ac.uk/
PIX analyses (HGMP)	http://www.hgmp.mrc.ac.uk/Bioinformatics/
PSORT II Prediction	http://psort.nibb.ac.jp/
SMART	http://smart.embl-heidelberg.de/

Chapter 3

Identification and characterisation of gene-trapped KRAB-ZFPs

3.1 Introduction

Despite the abundance of KRAB-ZFPs in the mouse and human proteomes, comparatively little is known about the specific biological functions and mechanism of repression employed by this protein family. Whilst at the start of this thesis, the precise functions of a number of KRAB-ZFPs and their target sequences had recently become known (Casademunt *et al.*, 1999; Skapek *et al.*, 2000; Zheng *et al.*, 2000), they were mainly thought of as a group of proteins involved in aspects of cell differentiation and proliferation that harboured a potent repressor activity. In humans, the target sequence of ZNF202 (a SCAN domain-containing KRAB-ZFP) suggested that this KRAB-ZFP is involved in regulating lipid metabolism (Wagner *et al.*, 2000). Similarly, ZBRK1 was known to repress *GADD45,* a gene involved in cell cycle regulation, by binding to a specific sequence within intron 3 (Zheng *et al.*, 2000). In rat, only one KRAB-ZFP (AJ18) target gene was known and was found to modulate the activity of *Runx2,* a master gene of osteogenic differentiation (Jheon *et al.*, 2001). In mice however, no KRAB-ZFP target genes had yet been discovered. Although comparative sequencing studies of KRAB-ZFP genes and the proteins they encode do exist (Mark *et al.*, 1999; Looman *et al.*, 2002; Shannon *et al.*, 2003), comprehensive functional studies are limited due to a lack of target genes and modes of analysis that incorporate the study of more than one protein. The gene-trap screen isolated 7 KRAB-ZFPs, which could be used as a starting point for such a comprehensive study. As a first step, I decided to identify and structurally characterise each gene-trapped KRAB protein. By doing so, I hoped to outline any similarities and differences within the candidate group in preparation for choosing group representatives for further functional studies.

The gene-trap approach employs a β-galactosidase-neomycin phosphotransferase (*β-geo*) reporter gene (lacking its own promoter and start codon), which more often integrates into the introns, rather than the exons of expressed genes. If the reporter gene is spliced in frame into the gene transcripts (Fig. 3.1A), translation results in a fusion protein that is expressed from the trapped genes endogenous

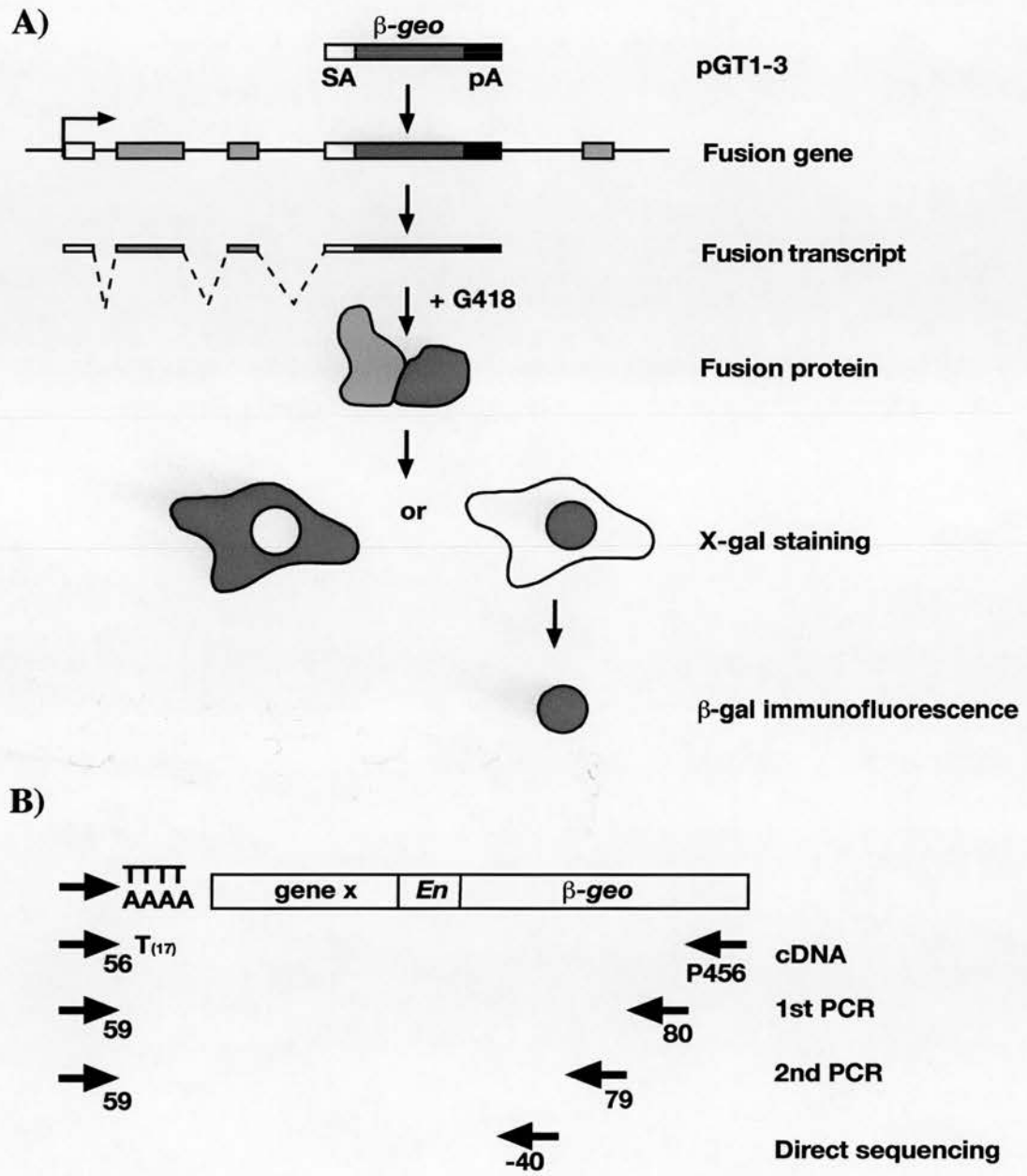


Figure 3.1 Gene-trap screening and sequencing strategy
A) pGT(1-3) containing the β-*geo* reporter gene is integrated into the intron of an expressed gene downstream of coding exons (grey boxes). The fusion gene is spliced in frame into the gene transcript via its splice acceptor (SA), taken from the mouse *engrailed 2* gene (*En2*). The polyadenylation (pA) signal ensures that the gene transcript terminates after β-*geo*. The transcript encodes a fusion protein possessing both neomycin phosphotransferase and β-galactosidase activities. Drug selection (G418) identifies cells that have integrated the vector into the introns of expressed genes in the correct orientation. X-Gal and immunofluorescence assesses the sub-cellular and sub-nuclear distribution. **B)** The sequence of trapped genes is established by 5' RACE. Products of the second PCR reaction are sequenced directly. Primers used are the same as those used by Sutherland *et al.*, 2001, and are described in section 2.3.2.

promoter, where the N-terminal amino acids are derived from the endogenous gene and the C-terminal amino acids derived from the amino acids are derived from the endogenous gene and the C-terminal amino acids derived from the *β-geo* cassette. X-Gal staining and immunofluorescence with an anti-β-gal (α-β-gal) antibody was used to identify ES or F9 cell clones with nuclear fusion proteins that may recapitulate the sub-nuclear localisation of the endogenous protein, if sequences that determine sub-nuclear compartmentalisation are N-terminal to the site of gene-trap integration (Sutherland *et al.*, 2001). 5' rapid amplification of cDNA ends (5' RACE), and sequencing of the products allowed a comparison of the sequences to genomic databases to be identified as known or novel (Fig. 3.1B). The 5' RACE protocol and the primers used were the same as those previously reported (Sutherland *et al.*, 2001) and are described in Table 2.3.

3.2 Characterisation of gene-trapped loci by FISH

Initial characterisation of trapped genes required their chromosomal locations to be identified because at the start of my PhD less KRAB-ZFP gene sequences and/or expressed sequence tags (ESTs) were available in the databases, and their assembly within the genome was either non-existent or in regions not yet sequenced. I took advantage of the *β-geo* cassette situated 3' to the endogenous gene sequence within the gene-trap lines, to determine the chromosomal locations of gene-trap integration sites by 2-D FISH analyses of metaphase spreads (section 2.9). Some KRAB-ZFPs are very similar to each other, and a single gene on one chromosome can often be related to multiple genes on different chromosomes (Bellefroid *et al.*, 1993). Therefore identification of chromosome location may also help to differentiate between highly similar gene/protein sequences, which would otherwise be difficult to distinguish. FISH can also rule out the possibility of more than one gene-trap integration per gene-trap cell line.

Probe DNA (pGT1-3) was labelled with biotin and hybridised to metaphase spreads from each gene-trap line (section 2.9). More than 20 metaphases from each gene-trap

line were analysed by Muriel Lee and karyotyped from the DAPI banding pattern of the chromosome on which the gene-trap signal was found. I then confirmed the chromosomal location of the gene-trap in each cell line by co-hybridisation of pGT1-3 probe with the relevant commercial mouse chromosome paint labelled with FITC (Fig. 3.2 and Table 3.1).

As expected, signal was detected on both chromatids of only one chromosome per gene-trap line, accounting for integration of the gene-traps at single genomic locations. Signal was sometimes detected on mouse chromosome 5 (MMU5) at the endogenous mouse *En2* locus (the only place in the mouse genome with sequence homology to part of the pGT1-3 vector, i.e. the splice acceptor was derived from the *En2* gene). The analyses suggest that for cell lines ES9 and F9/30A6, the gene-trap is on an abnormal chromosome in some of the cells. Although ES9 and F9/30A6 are situated on MMU13 and MMU9 respectively, most of the cells in these cell lines have abnormalities on these chromosomes, which may involve fusions with other chromosomes. These abnormalities may be a consequence of the gene-trap insertions or unrelated, but ultimately exclude the prospect of chimeric mice generation with the ES cell line. Strikingly, 5 of the trapped genes (ESKN205, ES261, ES492, ES510 and F9/30A6) appeared to map close to telomeres or centromeres, in agreement with the literature as human KRAB-ZFPs are known to have a biased distribution to telomeric, centromeric and/or fragile sites within the genome, which also seems to be the case in other tetrapods (Lichter *et al.*, 1992; Eichler *et al.*, 1998; Meiboom *et al.*, 2004).

3.3 Sequence identification of trapped genes

I repeated 5' RACE and direct sequencing for each KRAB containing gene-trap line previously reported (Sutherland *et al.*, 2001), plus one additional but as yet unpublished gene trapped in the same screen (ES cell line 492) in order to confirm previous RACE attempts and to obtain better quality sequence data (Table 3.1). Between 300~800 base pairs (bp) of 5' RACE sequence were obtained for all gene-trap lines (Fig. 3.3) and were compared with other sequences in the GenBank and Ensembl databases using the BLAST algorithm. The websites used for all

Figure 3.2 FISH analyses of gene-trap cell lines

MAA-fixed chromosomes from each mouse gene-trap line were hybridised with FITC-labelled commercial MMU paint (green) and biotin-labelled pGT1-3 probe (red). Grey-scale representations of the DAPI stained metaphase chromosome spreads from 3 gene-trap cell lines are shown.

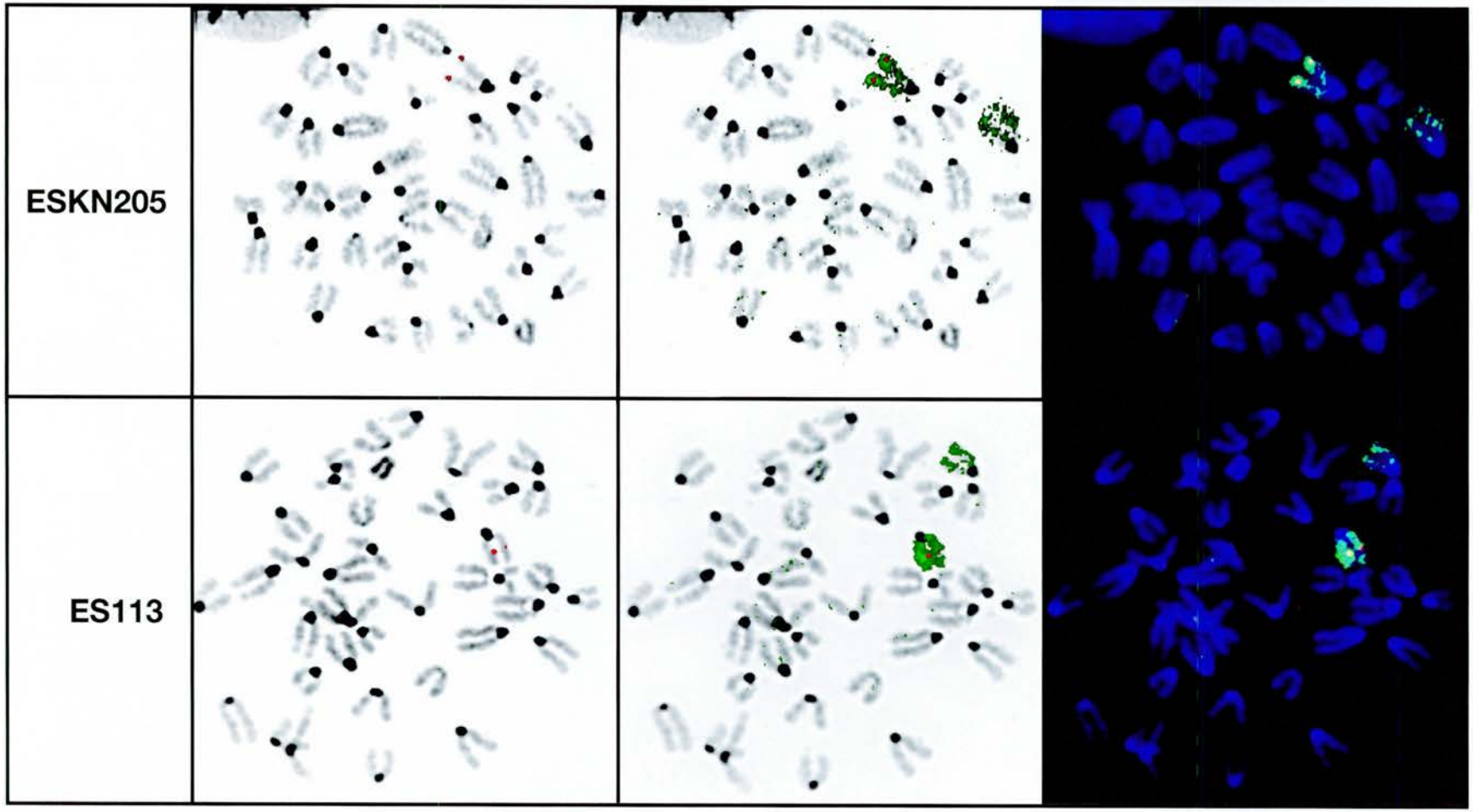


Table 3.1 Characterisation of KRAB-ZFPs

Cell line/gene name	*Mouse sequence (accession no./gene ID)*	*Chromosomal location in mouse*[I]		*Homologous sequence/region*[II]	*Protein domains*[III]	*Summary of genomic region*
		FISH	*Ensembl*			
ES9	MZF6D (AY149175)	13	13B3	Hs. none Rn. RNO14p22	KRAB A+C, *7x ZFs* 3xNLS (120, 175, 176)	highly repetitive genomic region. ~10 ZFPs span 1 Mb of surrounding region.
ES113	BC047396	13C1/C2	13 un	Hs. none Rn. RNO17p14	KRAB A+C, *? ZF*	poor genomic assembly of 13B3-C2. Possible rat homologue on RNO17p14.
ESKN205	BC058858, similar to *Rex2* (AF051348)	4E1/E2	4E1	Hs. none Rn. RNO5q36	KRAB A only	>5 ZFPs in surrounding 1 Mb. Duplicated paralogue of *Rex2*.
ES261	AK045713	16C3	16C4	Hs. HSA21q22.3 Rn. RNO11q12 Gg. GGA1, ~102.8 Mb	KRAB A+C only	exact genomic location unknown but region is highly conserved between species
ES492	hypothetical *Zfp647*, NM_172817	15E	15E1	Hs. *ZNF647*, NM_021061 (8q24.3) Rn. ENSRNOG00000004523	KRAB A+B, *13x ZFs.* NLS (219)	genomic region= AY187287. 3 ZFPs in surrounding 1 Mb region
ES510	L20450 (MUSDNABPZ/*Zfp97*)	17A	RACE on 17A2/3 L20450, un	Hs. un Rn. un	KRAB A+C, *13x ZFs.* NLS (367)	poor genomic assembly of 17A2/3. Chromosomal location of L20450 is un.
F9/30A6	similar to BC050776	9D2	un	Hs. none Rn. un	KRAB A(B?)	member of KRAB-ZFP cluster? Chromosomal location un.

[I] The chromosomal locations of trapped genes, as determined by genomic database searching at Ensembl and FISH. *Un*, unknown genomic location.
[II] Mammalian homologues of trapped genes or their possible orthologous clusters. *Un* , unknown genomic location of possible orthologue/s.
[III] Protein domains or motifs identified by database analysis are indicated in the order in which they appear in the protein sequence. Italicised domains are those removed from the trapped proteins by gene-trap integration. Numbers in brackets designate the amino acid position of the NLS start.

Figure 3.3 Alignments of 5' RACE sequences

Nucleotide alignments of each RACE sequence with their nearest BLAST/Ensembl hit were performed using the bl2seq algorithm at NCBI. Capital letters represent RACE sequence with no significant similarity to other sequences in the GenBank and Ensembl databases. Numbers represent nucleotide positions within the matching cDNA sequence and flank contiguous sections of similarity. Nucleotide matches with *En2* are boxed. Corresponding protein sequences of either the translated RACE sequence and/or the matching cDNA are shown in italics and the KRAB box is underlined. The translated protein sequence is reversed, but in the correct orientation. For ESKN205, the circled number designates the end of the match between ESKN205 RACE sequence and AF051348.

Es9

```
cttgtatgtcttcgagaactttgacatggttcttcaatgtgatggtcttggnatttgtaacctat RACE
cttgtatgtcttcgagaactttgacatggttcttcaatgtgatggtcttcccaattgtaacctat AK082760
  T  H  R  R  S  S  Q  C  P  E  E  I  H  H  D  Q  X  K  Y  G  I
299
agaagtaaggttccagtaggtctccaacataacatctttgtagagattcatctgggaaggatcca
agaagtaaggttccagtaggtctccaacataacatctttgtagagattcatctgggaaggatcca
 S  T  L  N  W  Y  T  E  L  M  V  D  K  Y  L  N  M  Q  S  P  D  L

gcaaattccattcttctttagtgaagttcacatgcacatcgtcaaaagtcactgcatccatgtct
gcaaattccattcttctttagtgaagttcacatgcacatcatcaaaagtcactgcatccatgtct
  L  N  W  E  E  K  T  F  N  V  H  V  D  D  F  T  V  A  D  M

aaatctcagagcttcctgaggtcccctcaatgcacccgacagcagatctcagagcaggacacaga
aaatctcagagcttcctgaggtcccctcaatgcacccgacagcagatctcagagcaggacacaga

gaaccattcaaccctgcacagccacagcgatcttctNAAAAAAAAAAAAAAAAAAANNNTTNTANG
gaaccattcaagcctgcacagccacagcgatcttct
                                    4
NNNNNCCCCCCCC
```

ES113

```
cttctgtgtcttttngaactttgacaatattcntcgatattaggggcttcccaattaaagcctat RACE
cttctgtgtcttctagaactttgacaatattcttcgatattatgggcttcccaattaaagcctat BC047396
R  R  H  R  R  S  S  Q  C  Y  E  E  I  N  H  A  E  W  N  F  G  I
271
agcantnangntcctgtaggtctccaccatgacatctttgtagagcttcttcngngaaggagnca
agcattgaggttcctgtaggtctccaccatgacatctttgtagagcttcttctgggaaggatcca
 A  N  L  N  R  Y  T  E  V  M  V  D  K  Y  L  K  K  Q  S  P  D  L

ncaacagcgccacacctcttgagtgaagttcacatgcaggtcctcataagtcactngcattcatg
gc-agagc-ccactcctcttgagtgaagttcacatgcaggtcctcataagtcact-gcattcatg
   L  A  W  E  E  Q  T  F  N  V  H  L  D  E  Y  T  V  A  N  M

tcgaacacgccgtagatcgctggcnaattccntaaggcccctggcaccaganattaaaacaaaan
tcg-ccacgtcgtagatcgctggc-aattccctaaggcccctggcagcagagattaaaac-aaat

cagaccaagcttCGCAGGNCTGCTCCGAAACGGGANAGGGACACANANCACCGGNACTAANCNNA
cagaccaagctt
          6
CCCNCTANNCCNCCCACTTCCCNCGGGCACTCGCCANGGNGNCANNGCTACNCCCTNNNNCACTA
CCGCCTCTTATCCNCCTNNCNACNNCTTCCNNCNACNATCTCNANATNNGCTCCNNNNCTCNCAC
TNTCNTAACANCCCNNNNNCGNCCCCCCCTTGNANTCNNNTCCNCNCCNTCACGCNNCNCNAGNC
CCCCCNATCCGNCATNCANCCNCCCTCANCNCNNNCAAGANANGNACCCGNCCCNNCTCNCNACN
GCNTAATCCNCATNCACGC
```

Figure 3.3 continued.

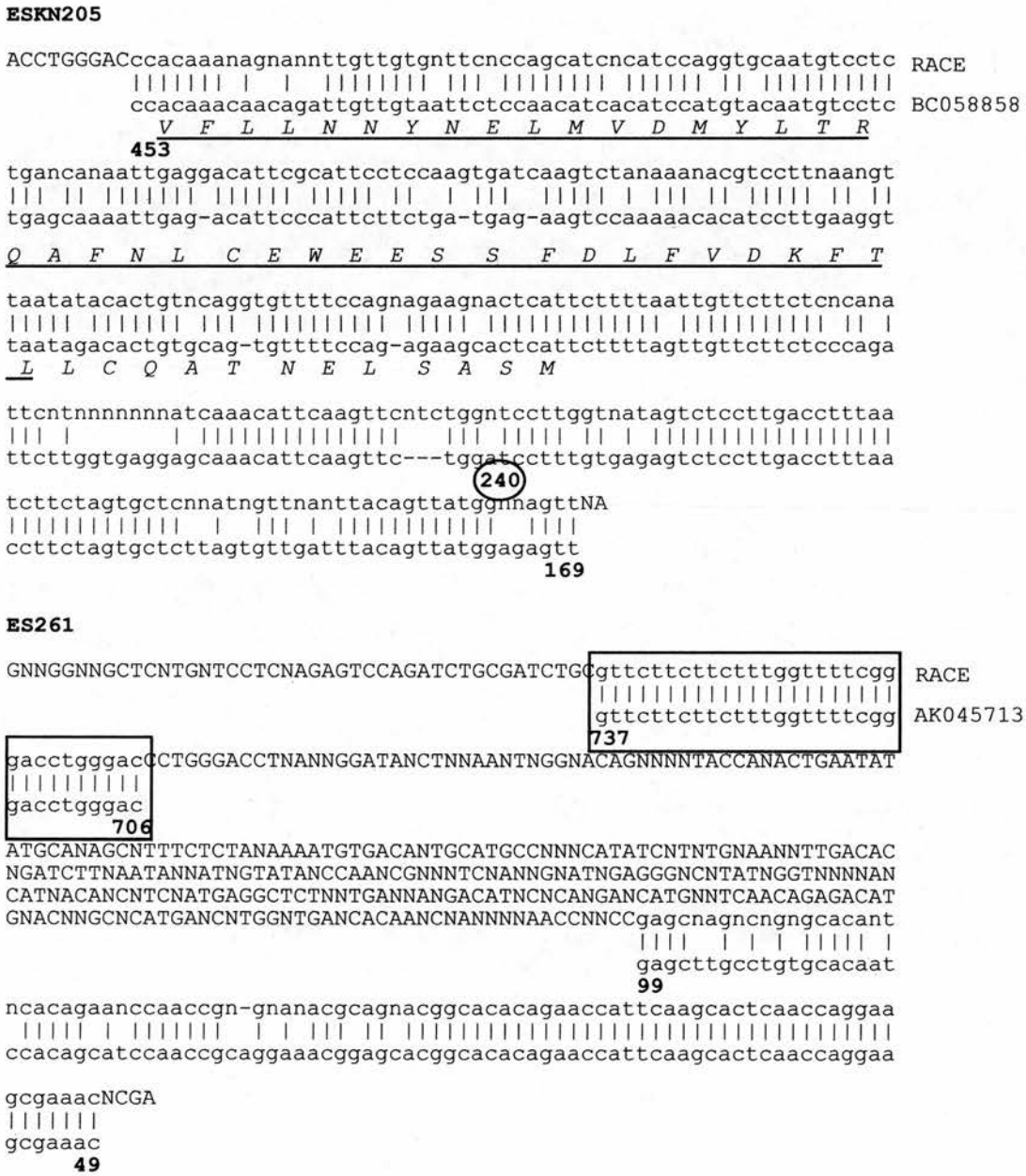
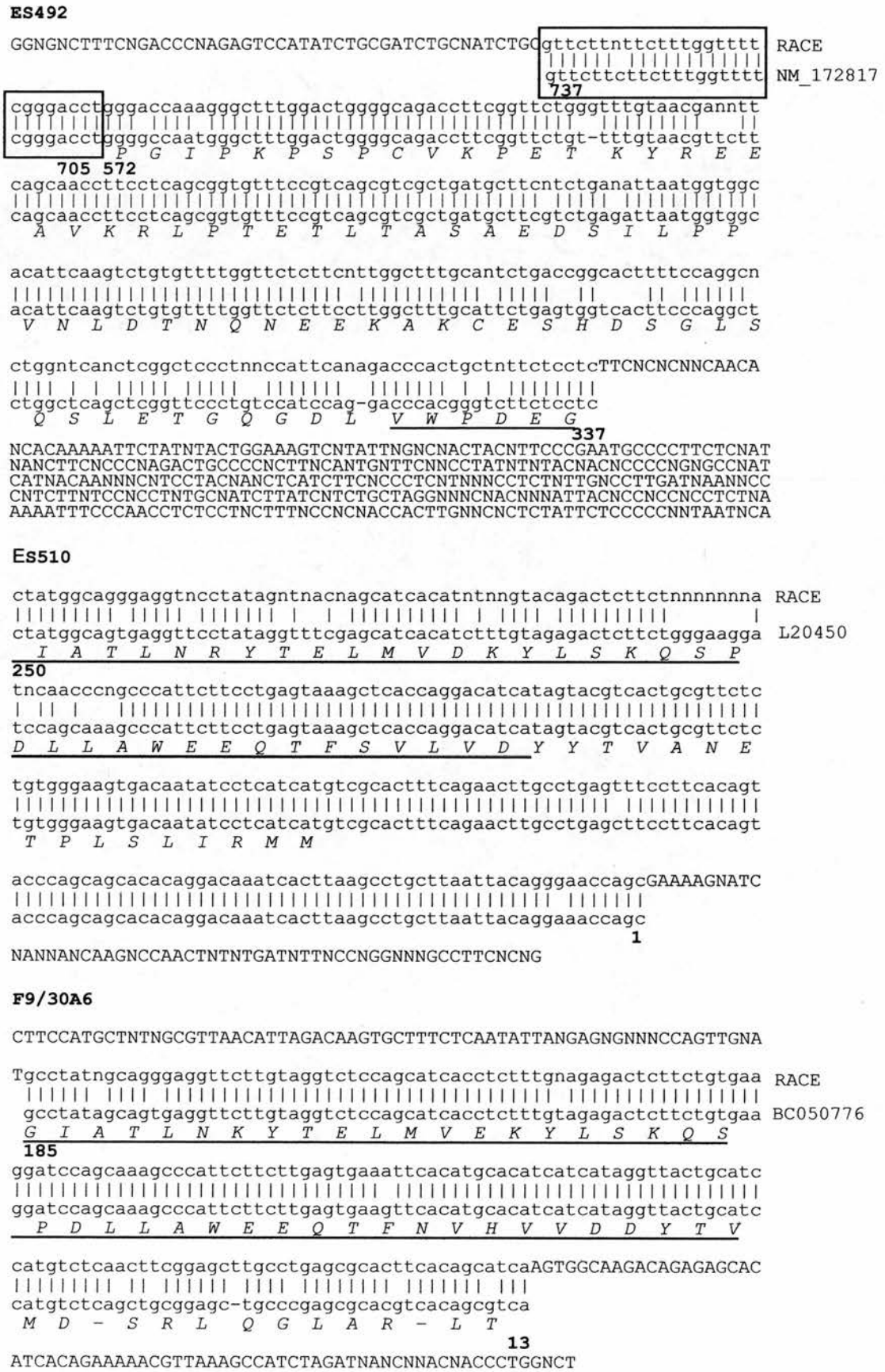


Figure 3.3 *continued.*



computational methods in this chapter and throughout the rest of the thesis can be mouse found in section 2.11. I also used the SSAHA (Sequence Search and Alignment by Hashing Algorithm) at Ensembl with pre-defined optimised parameter sets, to look for exact or 'almost exact' matches between the RACE sequences and the genomic assembly. SSAHA is a faster tool than BLASTN for matching and aligning DNA sequences with genomic assemblies. Predicted protein sequences for each trapped gene were assessed for conserved protein domains and motifs using InterProScan at EBI (European Bioinformatics Institute) and SMART (Simple Modular Architecture Research Tool). For gene-trap sequences with known gene and therefore protein sequences I used PSORT II to confirm that the nuclear localisation observed in the gene-trap lines is the same as predicted for the endogenous protein. PSORT II was also used to find the location of any nuclear localisation signals (NLSs) or other signal peptides. However, since all trapped genes were selected on the basis of their nuclear localisation, their NLSs must be present either before the site of gene-trap insertion (and therefore contained in the RACE sequence) or otherwise their localisation is achieved via transport along the endoplasmic reticulum (ER) and nuclear envelope (NE) and/or complexed to other NLS-containing proteins. The results of these searches are presented in Table 3.1 and as of the 25th October 2004, all data resulting from these searches is deemed to be correct. A more detailed description of each gene-trap loci follows.

3.3.1 ES9

BLASTN analysis indicates that the ES9 RACE sequence has 97% identity over 296 nucleotides to Riken cDNA, AK082760, which encodes a peptide with a KRAB domain (Fig. 3.3). Its protein sequence is only 95% identical to that encoded by the RACE sequence because of unknown nucleotides (Ns) in the RACE sequence that cannot be translated into amino acids (Fig. 3.3). ES9 and AK082760 are both 98% identical over 265 nucleotides to AY149175, a partial cDNA clone encoding known mouse KRAB-ZFP, MZF6D (Fig. 3.4A). The match is in the region encoding the KRAB domain and a portion of the 5' UTR. MZF6D has at least 7 ZF motifs and was originally isolated from a mouse testes cDNA library (Åbrink *et al.*, 2001) but has only recently been deposited in GenBank and published as a member of the

newly identified KRAB A+C family (Looman *et al.*, 2003). Despite various attempts, a full-length cDNA clone for this gene has not yet been isolated (Looman *et al.*, 2003). PSORT II predicts MZF6D to have a nuclear localisation and three NLS can be found in the ZF region (two of which overlap) (Fig. 3.4A). All three of the above sequences match to two overlapping Ensembl predicted genes within the same contig (AC101221.11.45569.258193) located at ~63.8 Mb on MMU13 in band B3 (Fig. 3.4B). This mapping data is consistent with that obtained by FISH analysis (section 3.2), although the ES9 cell line appears to be abnormal (data not shown). The match is in the KRAB encoding region of MZF6D/ES9, however of the two predicted genes (which share 95% nucleotide identity over this same region) only one encodes an identical KRAB domain (ENSMUSG00000058671, NM_183119), because an additional nucleotide in the remaining gene (ENSMUSG00000053212, Q8BKS1) after the start codon shifts the translational reading frame. Since the predicted genomic structures of neither gene seem to encode a full-length KRAB-ZFP (the KRAB domain of NM_183119 is immediately followed by a STOP codon), I used the two-sequence alignment algorithm at NCBI (bl2seq) to BLAST the corresponding genomic sequence of this contig (AC101221.11.45569.258193) with the exon sequences encoding MZF6D (5' UTR, KRAB A, KRAB C and ZF domains) (data not shown). The results indicate that the 5' UTR, KRAB A and KRAB C exons all match with 100% nucleotide identity to the genomic contig. Moreover, this genomic clone contains a duplication of each of these exons which share only ~96% nucleotide identity over the same length of sequence. Unfortunately, only 95% nucleotide identity is observed with the MZF6D ZF exon, suggesting that this exon is either not contained within this contig or that there are sequence discrepancies within either MZF6D and/or Ensembl. In fact, the ZF exon has over 20 hits within this contig and all have greater than 75% nucleotide identity. Collectively, this data implies that the DNA sequence contained in this genomic contig is highly repetitive. The orientation of MZF6D within the contig suggests that the ZF exon could be located in a contig upstream of this genomic clone, and this is not the case. Unorientated upstream contigs are also separated by gaps in the sequence and so the full genomic structure of this gene cannot be correctly identified. This analysis confirms that MZF6D is composed of at least three exons, the first of

which contains the 5' UTR and the start codon (ATG), the second and third encode the KRAB A and KRAB C domains respectively.

Within the 1 Mb region surrounding this locus I can find ~10 different ZFP genes, four of which also have a KRAB domain. The repetitive nature of ZF motifs and the fact that many ZFPs are duplicated in this region expose the problem of sequence assemblage in such regions and means that gene prediction programmes perform badly here. These other KRAB-ZFPs have very different KRAB domains compared to ES9/MZF6D and show only 40% to 50% protein homology over the full-length sequence (data not shown). Therefore, these other KRAB-ZFPs belong to a different subfamily of KRAB-ZFP genes in this region.

To identify if putative homologues in other species could help to decipher the genomic structure of ES9/MZF6D I searched for syntenic regions in human and rat. There is very little DNA sequence conservation between the mouse and human genome over a 1 Mb region surrounding ENSMUSG00000053212 at 63.8 Mb on MMU13 (data not shown) so I performed a SSAHA/BLASTN search of the human genome with the MZF6D nucleotide and protein sequence. I could not find an orthologous KRAB-ZFP gene or gene cluster in the human genome. The closest matching protein has ~48% identity over 269 amino acids confirming that some genes very distantly related to MZF6D/ES9 may exist in humans, as postulated by Looman *et al.* (2003). According to Ensembl, a break in synteny with the rat genome exists at ~63.9 Mb. MZF6D/ES9 lies ~100 kb upstream and is one of the last genes before the break. Although there are areas of sequence conservation between MMU13 at ~63.8 Mb with the telomere on the short arm of rat chromosome 14 (RNO14p22), no homologous MZF6D gene appears to exist in this region and neither does it contain any ZFPs of the Krüppel type (Fig. 3.4B). Post 63.9 Mb, synteny is observed with RNO7q22 and many genes are conserved in this region. A SSAHA/BLAST search using both the nucleotide and protein sequences of MZF6D indicates that MZF6D/ES9 is distantly related to numerous KRAB-ZFPs found on different rat chromosomes, supporting previous northern blot data by Looman *et al.* (2003). Most nucleotide and/or protein similarities (~61-68%) are found close to the

telomeres of RNO12q16, RNO3q43 and RNO1p13 (data not shown). However only RNO1p13 is a known ZFP cluster containing KRAB-ZFPs. Whether these genes are true orthologues remains to be proven.

3.3.2 ES113

By BLASTN, ES113 RACE sequence matched a RIKEN cDNA clone (BC047396) with 90% identity over 272 nucleotides (Fig. 3.3). BC047396 encodes a KRAB-ZFP with 7 ZFs but the match is in the 5' UTR and KRAB domain portion of the cDNA. ~60% of the mis-matching nucleotides in the RACE sequence are unknown nucleotides (Ns) therefore less amino acid identity is observed over this region (71%). A SSAHA search at Ensembl confirms that both the RACE sequence and BC047396 probably match genomic DNA fragments on MMU13 at around band B or C, consistent with FISH mapping of ES113 to MMU13 band C. However, the highest matches for both are in unorientated contigs that have not yet been assembled into the genome. There are many large gaps between the sequenced contigs in this region, probably due to assembly problems based on the repetitive nature of the genome here, since clusters of ZFP and KRAB-ZFP genes are interspersed throughout bands B3 to C2 on MMU13 (data not shown). Assigning a gene to ES113 is therefore not possible based on the RACE sequence alone. To determine whether rat and human orthologues of ES113 exist I performed BLAST/SSAHA searches on the respective genomes with ES113 RACE sequence and BC047396 cDNA. Ensembl does not detect a human orthologue with either the RACE sequence or BC047396, however MMU13 (B3-C2) does appear to have a highly conserved syntenic region in rat. A SSAHA search with ES113 RACE sequence shows hits over a 39 bp region to the short arm of rat chromosome 17 (RNO17p14). RNO17p14 is highly conserved with MMU13 band B3 and is also littered with KRAB-ZFP genes (I detected at least 9 different KRAB-ZFP genes over a 1 Mb region here). However, there are still many gaps in the rat sequence and some contigs are unorientated suggesting that a genuine rat orthologue may be mis-positioned. An SSAHA/BLASTN search with BC047396 reveals an 87% match to RNOXq35 over ~300 bp in the region of encoding the KRAB box. This match precedes subsequent hits (>6) to RNO17p14 over a similar region of the KRAB box

(data not shown). Whilst no portions of MMU13 between bands B and C are syntenic to RNOX, it may be that ES113 is orthologous to more than one rat gene as previously reported for other KRAB-ZFPs (Shannon *et al.*, 2003) or that this contig has been mis-assembled on RNOX. RNOXq35 is not associated with a KRAB-ZFP cluster in this region but rather a single, isolated KRAB-ZFP seems to exist here. In conclusion, the exact genomic location of an ES113 orthologue in rat cannot be determined on RNO17p14 since the genomic location of ES113 is unknown, however it is likely that one or more exists.

3.3.3 ESKN205

ESKN205 shares 80% (over 293 bp) and 81% (over 219 bp) nucleotide identity to full-length mouse cDNA, BC058858 and mRNA, AF051348 respectively (putative *Rex2* mRNA) (Fig. 3.3). Both of these genes encode KRAB domains only and are 93% identical to each other at the nucleotide level but differ at individual bases throughout the length of the sequence. Their KRAB boxes (also encoded in part by the RACE sequence) are 86% identical at the protein level. However the RACE sequence shows 83% nucleotide identity with the 5' UTR of BC058858 and this sequence is not contained in AF051348 (Fig. 3.3). Therefore BC058858 and AF051348 could be separate genes with a very similar KRAB domain. To verify this, I used the BLASTN algorithm at Ensembl to find out which part of the mouse genome the clones and RACE sequence map to. As expected, both of these sequences and the RACE map with the highest percentage identity and lowest e-value, to genomic DNA in band E1 on MMU4 consistent with FISH mapping of ESKN205. However BC058858 and AF051348 map on opposite strands of adjacent contigs (Fig. 3.5A). AF051348 matches to Ensembl predicted gene, ENSMUSG00000061969 at ~145.2 Mb. This gene is associated with ~10 transcripts, 8 of which encode a KRAB box in association with ~13 ZFs. The RACE sequence shows 73% protein identity to ENSMUSG00000061969 in the KRAB box region. BC058858 matches to Ensemble predicted gene, NM_145078, located on the opposite strand of the adjacent contig at ~145.5 Mb on MMU4 (Fig. 3.5A). Translation from the predicted start codon (ATG) in NM_145078 results in a putative protein, BAB27959.1 (putative *Rex2*) that contains the latter part of the

Figure 3.4 The sequence, genomic location and orthology of MZF6D/ES9

A) MZF6D partial cDNA sequence (lower case) with corresponding protein sequence (upper case). Bold text represents the start codon. The KRAB A and C domains (open and shaded boxes) are indicated and the ZF motifs are underlined. The gene-trap integration is located after arrowhead. A portion of the 5' UTR is shown preceding the start ATG and the potential NLSx3 in the ZF domain are shown in bold italics (the latter two overlap). **B)** Ensembl predicted genes (NM_183119 and Q8BKS1) are shown in their genomic region on MMU13 in band B3 (open boxes). Other KRAB-ZFPs in the surrounding region are circled. A multi-contig view of the rat syntenic region (RNO14p22) is shown and the amount of sequence conservation is indicated (pink/purple blocks).

KRAB domain and 5 ZF motifs. However this putative start codon does not have an identical match to the Kozak consensus sequence (A/GNNATGGG) (Kozak, 1986). On further inspection, I found that an ATG codon immediately upstream of the predicted start ATG results in a transcript that exactly matches BC058858. The standard genomic structure of a KRAB-ZFP gene normally spans 4 exons depending on the KRAB domain family to which it belongs (Shannon *et al.*, 1998 and references therein). The first exon generally contains the 5' UTR and/or the start ATG. The second and third usually contains the KRAB box (depending on whether an A, B/b or C box is present) while the last exon contains the ZFs. The Ensembl predicted transcript does not conform to this standard whilst the genomic structure of BC058858 does. I therefore conclude that the gene trapped by ESKN205 is NM_145087, and although this transcript encodes a KRAB only protein, the use of a second ATG codon further downstream can change the reading frame and result in a Krüppel type ZFP protein with truncated KRAB box. Hence, it is plausible to speculate that cryptic splices sites within this transcript could operate to join together these two separate functional domains. KRAB-only containing proteins have recently been identified as functional proteins with roles in transcriptional repression (Li *et al.*, 2003; Oh *et al.*, 2005). PSORTII predicts BC058858 to be localised to the nucleus and it has an NLS (KRHK) in the C-terminus that may be responsible for this. Gene-trap localisation data confirms this prediction (Chapter 4).

I counted ~5 other ZFPs in a 1 Mb region surrounding NM_145078, and checked for the presence of other KRAB-ZFP genes which may harbour a similar KRAB domain. Apart from the two genes above, a third gene, NM_198619 also contains a KRAB box. However it has only 36% amino acid identity with the RACE sequence and is clearly a different gene (data not shown). Although 100% nucleotide identity is not shared between the ESKN205 RACE sequence and NM_145087 (because of the quality of the RACE sequence), ESKN205 does share sequence identity with the 5' UTR of NM_145087 and not with ENSMUSG00000061969 (AF051348). The assembly in this region contains no unorientated contigs or gaps, therefore it is unlikely that ESKN205 is another gene very similar to the two described above. I therefore postulate that NM_145087 is likely to be ESKN205, which is a paralogous

gene to the AF051348 transcript (*Rex2*) and both belong to a larger sub-family of KRAB-ZFP genes within this 1 Mb region. NM_145087 is likely to be the result of a duplication and subsequent inversion of the full-length KRAB-ZFP gene, ENSMUSG00000061969, which is associated with many *Rex2* type transcripts including AF051348.

According to Ensemble, there is little sequence conservation between MMU4 at ~145 Mb and the human syntenic region. However, there is strong sequence conservation with RNO5q36 (Fig. 3.5A). A predicted rat KRAB-ZFP gene (ENSRNOG000000247078) shares ~70% amino acid identity with the KRAB domain of both ENSMUSG00000061969 and NM_145087 whilst ~48% identity is observed over the entire coding region. Consequently, an orthologous rat gene does exist but it is predicted to encode up 16 ZF motifs depending on the transcript. This suggests that it is an orthologue of ENSMUSG00000061969 since different transcripts of this gene also encode many ZF motifs. In support of this hypothesis, both of these genes seem to be of a proportionate length and span the same length of genomic sequence in either genome. In light of these observations I could not find a duplicated copy of the rat orthologue in its immediate vicinity, supporting the hypothesis that the duplication of NM_145087 is mouse specific, although the rat assembly does contain gaps in this region.

3.3.4 ES261

ES261 RACE sequence matches with 83% (over 91 bp) and 87% (over 79 bp) nucleotide identity to partial cDNAs AK045713 and AK021061 respectively, which are both 100% identical to each other over a 91 bp region and described as being similar to the KRAB-ZFP Krüppel-ZFP F80-L (Fig. 3.3). Most mis-matches in the aligned region can be explained by Ns in the RACE sequence and the aligned sections are separated by ~300 bp of sequence for which the RACE sequence quality is extremely poor. According to Ensembl, the matching portions of AK045713 and AK021061 map to exactly the same genomic location at the very tip of MMU16 in band C4 at ~98.8 Mb, although recently the full AK045713 cDNA was found in an unassembled contig. The RACE sequence also BLASTS to this region and FISH

analyses confirm gene-trap integration into the telomeric region of MMU16. The RACE sequence matches both of these cDNAs in the region upstream of the KRAB-domain corresponding to the 5' UTR. However, the ES261 RACE sequence is of poor quality in the likely KRAB encoding region because it contains ~21% unknown nucleotides. A frameshift mutation also occurs in the RACE sequence thus direct translation results in protein sequence, which does not match the putative protein encoded by AK045713 with any similarity. Since matching sequence is in the 5' UTR, one cannot compare the amino acid similarity. The start codon of AK045713 is after the match with the RACE sequence. Moreover, the product of translated AK045713 appears to be a truncated KRAB-ZFP, containing the KRAB domain only but no ZF motifs. An in frame stop codon after the first 64 amino acids of AK045713 suggests that this cDNA could be an example of a KRAB domain-only containing protein rather than an alternatively spliced isoform. PSORTII predicts AK045713 to be a nuclear protein in agreement with the gene-trap localisation data (Chapter 4).

Only three ZFP genes (none of which contain a KRAB domain) are found within a 1 Mb from the telomere of MMU16, all of which have orthologues on RNO11q12. Orthologous genes are also observed in the human (HSA21q22.3) and chicken genomes (GGA1, ~102.8 Mb) (data not shown). However, gaps in the mouse genome at this location and unassembled contigs are preventing the ES261 orthologous gene from being identified and I could find no KRAB-ZFP genes in the same region in either of the other genomes.

3.3.5 ES492

I compared the ES492 RACE sequence with all sequences available at NCBI GenBank and Ensemble. This revealed 94% nucleotide identity with hypothetical mouse mRNA, NM_172817 (Fig. 3.3), which encodes the KRAB-ZFP, ZFP647 (6030449J21). The translated RACE sequence has only 65% identity with Zfp647, because of the number of Ns in the sequence but no other KRAB-ZFP in the mouse genome showed similarity with the RACE sequence in a TBLASTN search (data not shown). A genomic clone (AY187287) located in band E1 on MMU15 was found

to contain NM_172817, which spans ~7.69 kb of genomic DNA (between 78.13 and 78.14 Mb) on MMU15 (Fig. 3.5B) consistent with FISH mapping. Its predicted protein coding sequence constitutes 4 exons separated by 3 introns, as expected of a KRAB-ZFP with more than one KRAB box (Fig. 3.5B). An exon containing only 5' UTR is found upstream. There are two ZFPs ~200-300 kb either side of this gene, neither of which contains a KRAB domain although a KRAB-ZFP is found ~500 kb downstream. This region of MMU15 is highly conserved on RNO7q34 and an orthologous gene exists (ENSRNOG0000004523). In humans, the orthologous gene (*ZNF647*, NM_021061) is at the telomere of HSA8q24.3, just before a break in synteny with the mouse genome (gene search, NCBI). The most recent version of Ensembl does not recognise this homology although previous releases have and have annotated the human orthologous gene (data not shown). Orthologous genes in all species above encode a highly conserved KRAB-ZFP with a KRAB box and 13 ZF motifs. 81% and 96% amino acid identity is found between the human and mouse, and rat and mouse proteins respectively (Fig. 3.6). This conservation in amino acid sequence highlights a possible important function for this protein. Such high levels of similarity are often seen between orthologous KRAB-ZFPs (Jheon *et al.*, 2002). In the case of ES492, the KRAB and ZF domains of mouse and rat are 100% identical. In humans 89% and 98% identity is found between the KRAB and ZF domains respectively. The linker regions are less conserved with 86% identity between mouse and rat orthologues whilst an additional 21 amino acids are found in the human protein. The similarity in the ZF region implies that this protein in all 3 species regulates the same genes.

3.3.6 ES510

ES510 has 90% nucleotide identity (Fig. 3.3) and 76% amino acid identity with known gene L20450 (MUSDNABPZ/Zfp97), as previously confirmed (Sutherland *et al.*, 2001). Zfp97 is a KRAB-ZFP with 13 ZFs and was first identified as a novel cDNA expressed in neuroblastoma cells (Wick *et al.*, 1995). Ensembl identifies L20450 to be located in a currently unknown contig, which is not yet sequenced, but is probably located somewhere on MMU17. The RACE sequence is situated in a contig known to be located on MMU17 on the boundary between band A2 and A3 (a

Figure 3.5 Genomic location and orthology of ESKN205 and ES492
A) Ensembl predicted genes (NM_145087 and ENSMUSG00000061969) are shown on MMU4, band E1 (open boxes). A multi-contig view of the rat syntenic region (RNO5q36) is shown and the orthologous rat gene is circled. **B)** Ensembl predicted gene, NM_172817, is located on MMU15, band E1 at ~78.14 Mb (open box). The transcript encodes 4 exons (numbered 1-4), with the largest containing the ZFs (4). The syntenic region and orthologous gene in rat is indicated by a corresponding black line. ZFPs are circled and the KRAB-ZFP is indicated by a star.

A)ESKN205

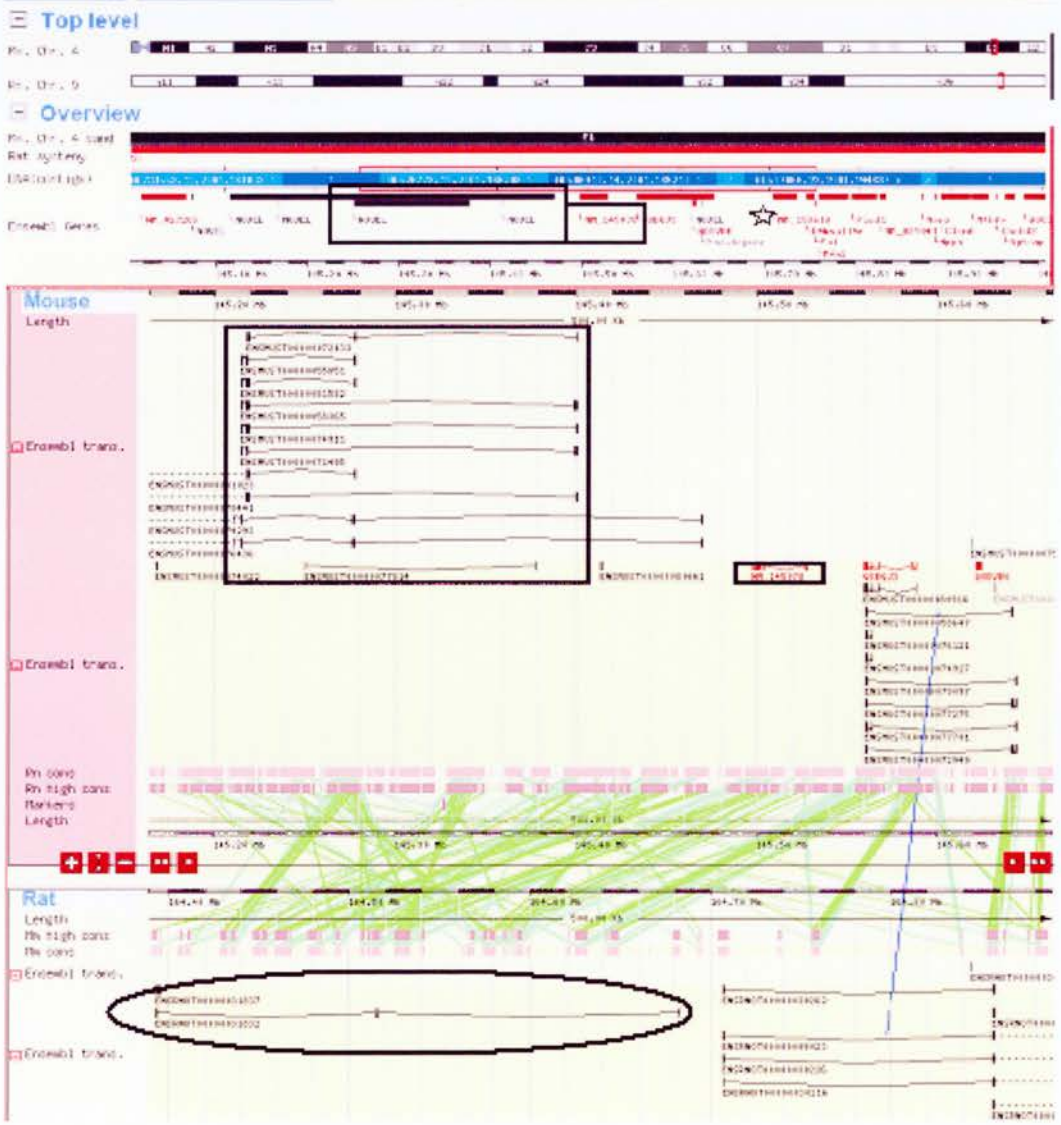
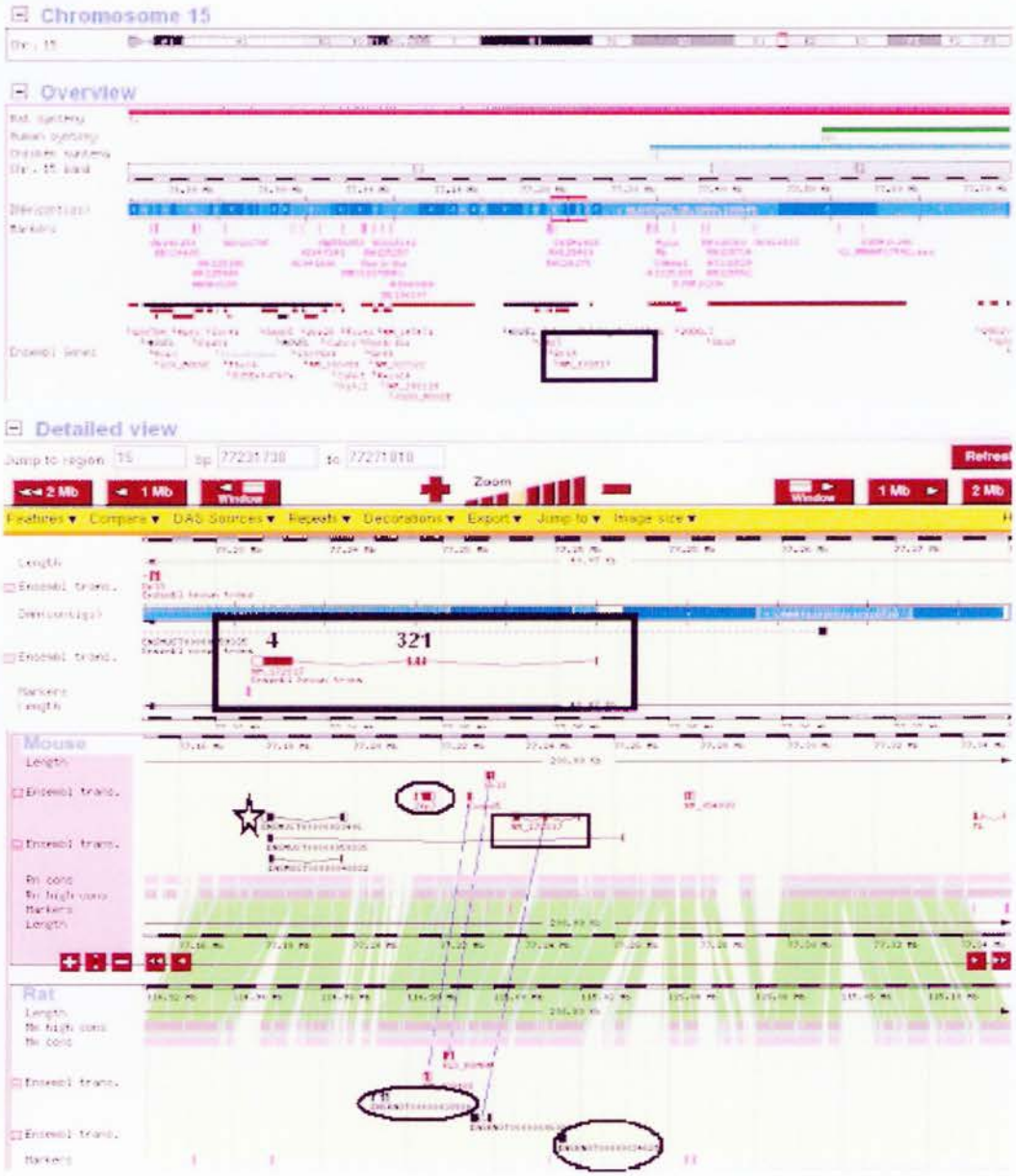


Figure 3.5 *continued.*

B) ES492



region just below the centromere), in accordance FISH mapping. However, although the assembly in this region is very poor and contains many gaps, a 4 kb region with high sequence conservation to RNO1q12 can be found (data not shown). At this same region there is a break in synteny with the human genome, where a syntenic boundary between HSA5 and HSA19 meets. I BLASTed L20450 at Ensembl and searched for possible orthologous genes in the rat and human genomes. However, since the genomic location of L20450 is unknown, no direct orthologues could be identified in either species. The best protein match in humans is ZNF442 and is found on HSA19p13.2. ZNF442 (NM_152601) has 52% amino acid identity with Zfp97 and is a KRAB-ZFP with 16 ZFs. ZNF422 is situated in a ZFP cluster on HSA19p13.2, a chromosome that contains a disproportionately large fraction (~1/3) of human KRAB-ZFPs, and this region is proposed to be rapidly expanding (Shannon *et al.*, 2003). The mouse syntenic region is unknown, because this part of the human genome is associated with a break in synteny with the mouse genome. Therefore ES510/Zfp97 appears to be mouse specific. PSORTII also predicts L20450 to be a nuclear protein.

3.3.7 F9/30A6

BLASTN identified that F9/30A6 is 93% identical at the nucleotide level to an mRNA expressed in brain, BC050776, but this identity is within the KRAB encoding region only (Fig. 3.3). The sequence upstream of the KRAB box lacks any identity with the RACE sequence. At the protein level, the KRAB box of BC050776 is 92% identical to the translated RACE sequence enforcing that these two genes may at least share the same KRAB encoding exon. A BLASTN search at Ensembl demonstrates that both the RACE sequence and BC050776 show similar alignments to regions scattered on at least 10 different chromosomes (data not shown). Although none of these regions map to the telomere of MMU9, which is the site of gene-trap integration determined by FISH mapping (section 3.2); it is worth noting that the highest scoring hits are Ensembl predicted transcripts, which are located in contigs that have not yet been assembled in the mouse genomic database. Therefore, Ensembl cannot predict the genomic location of F9/30A6. Sequences with unknown genomic locations are often associated with portions of the genome that contain

highly repetitive sequences, making individual genomic fragments difficult to consign. This may be the case for the above, suggesting that F9/30A6 may be a member of a KRAB-ZFP cluster and/or be paralogous to other mouse KRAB-ZFP genes. BC050776 encodes a KRAB containing protein with no ZF motifs due to an in frame termination codon after the KRAB domain. It is therefore possible that F9/30A6 also lacks ZFs. I searched for orthologous sequences within the rat and human genomes and in contrast with the human genome; at least 10 different rat chromosomes produced nucleotide alignments with both BC050776 and the RACE sequence. The best-hit rat protein is an Ensembl predicted protein that has 67% identity to BC050776, but this is based solely on the KRAB domain, a region which often shows high levels of protein similarity even though they are not orthologous (Bellefroid *et al.*, 1991). In human, the highest scoring nucleotide matches were on HSA19p13.2 but these hits are only found when the filters are removed. No nucleotide matches are obtained with the standard Ensembl filter sets. A BLASTP search of the human genome at Ensembl using the standard filter sets shows the best protein match is also present on HSA19p13.2 but is only ~58% identical to translated BC050776. Taken together I conclude that no orthologous gene exists in humans and since I do not know the precise genomic location of F9/30A6 in mouse, a specific rat homologue can neither be assigned, although it is likely one exists.

In summary, I have described the identification of 2 previously identified (ES9 and ES510) and 5 novel KRAB domain-containing genes. Of these novel genes, 3 are located in known genomic locations (ESKN205, ES261, ES492). Surprisingly, 3 genes are found to be most similar to KRAB-only containing proteins. So far, only 2 KRAB-only genes have been published and have functional roles. Although the ES113 and F9/30A6 genes could not be identified during the course of this thesis, significant improvements in the genomic databases over three years has made their initial study possible. The gene-trap has not preferentially integrated into any particular chromosome since out of 7 gene-trap lines, 6 locate to different chromosomes. However, although the actual genomic structure of most genes is unknown, there does seem to be a preference as to where the gene-trap integrates within the gene. In all cases, integration is after the KRAB domain, as each RACE

sequence contains either the 5' UTR and/or the KRAB box. Presumably the vector inserts into the intron separating the KRAB and the linker region/ZF coding exons. This might be expected since the ZF are encoded in a single exon and this intron is generally the largest in size when compared to the introns separating the 5' UTR and KRAB containing exons (Phillipe Gautier, Pers. Com.). Less likely is the possibility that the DNA in this region has a particular conformation that allows better access for the gene-trap vector. There is no preference for integration into sites of individual or clustered KRAB-ZFP genes and both are represented in this analysis, but the preferred distribution of KRAB-ZFPs in particular genomic regions (Lichter *et al.*, 1992) is demonstrated by the fact that 5 of the trapped genes are close to either telomeric or centromeric regions. It is interesting to note that some of the murine KRAB-ZFP genes in this study are orthologous to either rat genes alone (ES9, ES113, ESKN205) or both human and rat genes (ES261, ES492), suggesting that some of the gene-trap KRAB proteins are conserved between species and may therefore have fundamental functions.

3.4 Characterisation of KRAB sub-families

KRAB-ZFPs are divided into 4 sub-families on the basis of their KRAB domain. The newly identified KRAB C domain, once thought to be part of the linker region, has recently been described and is always found in conjunction with a KRAB A box. The 21 amino acid KRAB C domain is encoded by a separate exon and appears to strengthen the interaction between the KRAB domain and KAP-1 as indicated by yeast two-hybrid experiments (Åbrink *et al.*, 2001; Looman *et al.*, 2003), although it does not seem to affect the potency of transcriptional repression as indicated by repression assays with GAL4 DBD-KRAB fusion proteins in mammalian cells (Åbrink *et al.*, 2001). Therefore, the KRAB C domain may have physiological implications for transcriptional repression by KRAB-ZFPs (Looman *et al.*, 2003).

I used the CLUSTALW algorithm at EBI to produce alignments of the KRAB domains found in the gene-trap screen. I firstly aligned all the KRAB A domains encoded by either the translated gene-trap sequence or its closest matching gene, with the known KRAB A domains of 13 mouse KRAB-ZFPs used in an initial study

Figure 3.6 Protein alignment of KRAB-ZFP 492 homologues
The CLUSTALW algorithm was used to align KRAB-ZFP 492 protein sequences from human (NM_021061), mouse (NM_172817) and rat (ENSRNOG0000004523). Identical amino acids are highlighted. The KRAB domain is boxed and the ZF motifs are underlined. Red text indicates the sequences used to make antibodies. Bold italics represent antigenic regions in the linker sequence as predicted by PIX analyses.

by Mark *et al.* (1999) (Fig. 3.7A). ES510/Zfp97/DNABPZ, ES113, ES9, F9/30A6 and ES261 were found to be members of the KRAB A family. Those proteins with an A box similar to those observed in conjunction with a B box (ESKN205 and ES492) were further aligned with known B and b boxes from 8 mouse KRAB-ZFPs also used in the initial study (Fig. 3.7B). This shows ES492 to be a member of the KRAB A+B family. ESKN205 was found to lack a B box, even though it aligns within the A+B family. This has been reported for other KRAB-ZFPs (Mark *et al.*, 1999) and suggests that the B box may either not exist, or be differentially spliced, in this isoform. I favour the former since in this case some similarity can be seen with the highly divergent B box, but it seems that ESKN205 has further diverged from this consensus and this sequence may be a remnant of a previous B/b box. With the KRAB A family members, I aligned the 21 amino acid region subsequent to their KRAB A box with 14 known KRAB C containing mouse KRAB-ZFPs, identified in a recent study by Looman *et al.*, (2004). ES9 (MZF6D) and ES510 (Zfp97/MUSDNABPZ) were already known to contain a KRAB C domain and had been previously aligned in this study. Out of the 3 remaining sequences, ES113 and ES261 were also found to contain KRAB C boxes (Fig 3.7C). The protein sequence of F9/30A6 deteriorates after the KRAB A domain preventing further analysis with the C box although the RACE sequence suggests that a 21 amino acid long domain may be present.

Of the possible 4 KRAB-ZFP families, 3 are represented in this gene-trap panel (A, A+C, A+B). ESKN205 appears to be a KRAB A+B family member without a B box. A few residues were conserved with the b box so ESKN205 may represent a KRAB-ZFP whose b box has rapidly diverged. The KRAB A+B family is thought to be the origin of both the KRAB A+b and KRAB A families (Looman *et al.*, 2002). This is based on the fact that all human members of the KRAB A+b family are located in one single cluster on chromosome 19 (ZNF45 family; Dehal *et al.*, 2001) and some genes encoding KRAB A-ZFPs completely lack the B/b encoding exon (Vissing *et al.*, 1995; Mark *et al.*, 1999). The discovery of the KRAB C box and the results presented here indicates that this may not be the case. Since all human KRAB C containing KRAB-ZFPs are found on chromosome 19, one might be tempted to

speculate that the KRAB C box diverged subsequent to the KRAB A+ B family, as is the case with the A+b family. However, in mice, KRAB C containing proteins are found on chromosomes 7, 17, 11, 13 and 16 (Looman *et al.*, 2004 and this thesis) and in both isolated KRAB-ZFPs (ES261) and in ZFP clusters (ES113) further supporting the KRAB A+C family as a founder family. The only two KRAB-ZFPs found in vertebrates other than mammals are Xenopus Xfin and chicken cKr1. They both have KRAB domains that fall into the A+B family based on comparative alignments of their ZF regions with those of mouse and human (Looman *et al.*, 2002). In this study, the relationship between each KRAB sub-family is the same regardless of the region used to perform alignments (Looman *et al.*, 2002). Whether a domain similar to the KRAB C box exists in these vertebrates remains to be determined. The fact that 4 of the KRAB-ZFPs (possibly 5) contain a KRAB C box may signify an as yet uncharacterised importance for this domain. Whilst the KRAB C is not obviously required for gene repression its high conservation suggests it may be responsible for vital protein-protein interactions that remain to be identified. The presence of a KRAB A and B box in ES492 suggests that this protein does not belong to any larger sub-family of KRAB-ZFP genes and is possibly an ancestral gene. In support of this, only one other KRAB-ZFP gene possessing a significantly different KRAB domain is found ~500 kb downstream. ES261 and ES492 would both be postulated to be ancestral genes since they are in orthologous regions in human and rat.

3.5 Cloning and structural analyses of KRAB-ZFP 492

Since strong identity was found between the ES492 gene-trap RACE sequence and hypothetical *Zfp647*, I decided to sub-clone and sequence (section 2.2.13) the cDNA encoding KRAB-ZFP 492 (EST BI656339, IMAGE clone 5326813) to ensure that the whole coding sequence of the gene was present in the IMAGE clone. The primers used for all sequencing reactions of ES492 are shown in figure 3.8B (arrows) and these have been used for all confirmations of 492 sequences throughout this thesis. By comparison of the sequences from both mouse genomic DNA and the

A)

```
consensus A              ESVTFRDVAVDFSQEEWQLLDPAQRNLYRDVMLENYRNLVSL (human+mouse)
(Mark et al., 1999)
DNABPZ/ES510          -NAVTYYDVLVSFTOEEWALLDPSOKSLYKDVMLETYRNLTAI 42
ES113                 -NAVTYEDLHVNFTOEEWALLDPSOKKLYKDVMVETYRNLNAI 42
F9/30A6               -DAVTYDDVHVNFTQEEWALLDPSOKSLXKEVMLETYKNLPXI 42
MZF31                  DIVTYDDVHVNFTWEEWALLDPSOKDLYRDVMLETYRNLAAI 43
ES261                  DVLTHDDVHVNFTREEWALLNPSOKSLYKDVMLETYRNLNAI 43
MZF13                  DAVTYEDVHVNFSREEWVLLDPSOKSLYKDVMLETYWNLTCI 43
ES9                   -DAVTFDDVHVNFTKEEWNLLDPSOMNLYKDVMLETYWNLTSI 42
pMLZ-8                -EPVTFEDVAVNFTSGEWTLLDSSQKKLYRDVMKENFLNLISI 42

Zfp93                 -EMVTFRDVAVVFSEEELGLLDAAORKLYHDVMLENFRNLLAV 42
Zfp94                 --MVTFRDVAVVFSEEELGLLDAAORKLYHDVMLENFRMLLSV 41
pMTZ-1                -EAVTFKDVAVVFTKEEFRLLDSAORTLYQDVMVENFRNLLSV 42
Zik1                  -GCVTFQDVAICFSHEEWRLLDETQRLLYLSVMLQNFALINSQ 42

NK10                  -ESVTFKDVAVNFTOEEWHHVGPAORSLYRDVMLENYNHLVSL 41
ES492                 -AKVTFEDVAVLLSQEEWARLGPAORGLYRHVMMETYGNVVSL 42
MUSKID1               -VSLTFEDVAVLFTRDEWKKLVPSQRSLYREVMLENYSNLASL 42
Zfp30                 -STVMFRDVAVGFSQEEWECLSAYERDLYRDVMLENYSHLVSL 42
MZF22                 -EMLSFRDVAIDFSAEEWECLEPAOWNLYRDVMLENYSHLVFL 42
ESKN205               ---LTFKDVFLDFSSEEWECLNFAORTLYMDVMLENYNNLLFV 40
Zfp-37                -MATSEPAESDAVRAKEWEQLEPVQRDVYKDTKLENCSNPASM 38
```

B)

```
consensus B                    GLQVSKPDLITKLEQGEEPWIVKREIARATSP  (human+mouse)
(Mark et al., 1999)
NK10                           GYQVSKPEVIFKLEOGEEPWISEKEIQRPFCP- 32
ES492                          GLPGSKPVVISQLERGEDPWVLD---------- 23
Zfp30                          GCSISKPDVITLLEOGKEPWMIVRAEKRRWSR- 32
MZF22                          GLASCKPYLVTFLEQRQEPSVVKRPAAATVHP- 32
consensus b                    GHQLFKHDXISQLEREEKLWMMKXATQRGDSS  (human+mouse)

Zfp93                          GCQ-SPNKMAPLDTTGIRCLPLGQLPCWQMTS- 31
Zfp94                          GDK-NPEEMESLEEVGLRHLSHEALFCSQIW-- 30
Zfp-37                         GNODPKQDIVSVLEEEEPSSGKGKKASPSSLK- 32
pMTZ-1                         EYQLFKRD-KPYLEREEKPQMRRAAPRERDSGI 32
Zik1                           GCGHKTEDEERRVSTRASKGLRSETTPKTN--- 30

ESKN205                        ----ENHCICGNYEKEKVLGQDTQH-------- 21
```

C)

```
consensus C                                  GKKWKDQNIEEEYQNPRRNLR  (human+mouse)
(Looman et al., 2004)                        GYNWEDHNIEEHCQNSRRNGR  (mouse)

XM_145549                                    GYNWKDHNIEEHSONDRRNGR- 21
XM_142469                                    GYNTEEHNIEEHSONDRRYGR- 21
XM_145552                                    GYKWKDHNVGEHSONDRRYGR- 21
XM_142615                                    GYNWKDHNIEEHCONDRRYGR- 21
XM_205635                                    GYNWKDHQNEEHCONDRRYGR- 21
MZF13                                        GYKWKDHDNEEYCONSGRHGR- 21
6D                                           GYKWEDHHIEEPCOSSRRHTR- 21
ES9                                          GYKXQDHHIEEPCOSSRRHT-- 20
ES113                                        GFNWEAHNIEEYCOSSRRHRRC 22
ES261                                        GYNWEDNNIEEQCQSSRRNGR- 21
```

Figure 3.7 CLUSTALW alignments of gene-trap KRAB domains
CLUSTALW alignments of the KRAB A **(A)**, KRAB B **(B)** and KRAB C **(C)** domains were performed. The consensus sequence of each KRAB family is shown (taken from Mark *et al.*, 1999) and conserved residues matching this consensus are highlighted.

A)

B)

```
1                                                GCCCCACAGGATCGCCTGATTTCTGTTGTCCTCGCAGTCTTGAGATACCAAGatggcagcagcc
1                                                                                          M  A  A  A
65    gggcttctaccgctgccagcagcaccacaggccaaagtgacctttgaggatgtggctgttctgctgtctcaggaggagtgggcc
5      G  L  L  P  L  P  A  A  P  Q  A  K  V  T  F  E  D  V  A  V  L  L  S  Q  E  E  W  A

149   cgcctgggccctgctcagcgggcctctaccgacacgtcatgatggagacctacgggaatgtagtctcactgggacttccaqqa
33     R  L  G  P  A  Q  R  G  L  Y  R  H  V  M  M  E  T  Y  G  N  V  V  S  L  G  L  P  G

233   tctaaqcctgttgtgatctcccaqctqqaqcqaqqaqaaqacccgtqqgtcctggatggacagggaaacgagctgagccagagc
61     S  K  P  V  V  I  S  Q  L  E  R  G  E  D  P  W  V  L  D  G  Q  G  T  E  L  S  Q  S

317   ctgggaagtgaccactcagaatgcaaagccaaggaagagaaccaaaacacagacttgaatgtgccaccattaatctcagacgaa
89     L  G  S  D  H  S  E  C  K  A  K  E  E  N  Q  N  T  D  L  N  V  P  P  L  I  S  D  E

401   gcatcagcgacgctgacggaaacaccgctgaggaaggttgctgaagaacgttacaaaacagaaccgaaggtctgccccagtcca
117    A  S  A  T  L  T  E  T  P  L  R  K  V  A  E  E  R  Y  K  T  E  P  K  V  C  P  S  P

485   aagcccattggccccagattgcccacgggttgaaccctagtgtacccgttgcccgaccccagacggctcccagtgttgagaga
145    K  P  I  G  P  Q  N  A  H  G  L  N  P  S  V  P  V  A  R  P  Q  T  A  P  S  V  E  R

569   ccctatatctgcattgagtgtggaaagtgctttggtcggagctcccatctccttcagcatcagcgaatacacactggcgagaag
173    P  Y  I  C  I  E  C  G  K  C  F  G  R  S  S  H  L  L  Q  H  Q  R  I  H  T  G  E  K

653   ccctacgtatgccatgtatgcggaaaggctttcagccagagttccgtcctgagcaagcacaggcggatccacacgggcgagaag
201    P  Y  V  C  H  V  C  G  K  A  F  S  Q  S  S  V  L  S  K  H  R  R  I  H  T  G  E  K

737   ccctacgagtgtaacgaatgtgggaaagcctttcgcgtgagttccgaccttgcccagcaccacaagatccacaccggagagaag
229    P  Y  E  C  N  E  C  G  K  A  F  R  V  S  S  D  L  A  Q  H  H  K  I  H  T  G  E  K

821   cctcacgagtgtctggaatgtgggaaggcgttcacccagctctcccacctcatccaacatcagcggatccacacgggggagcgg
257    P  H  E  C  L  E  C  G  K  A  F  T  Q  L  S  H  L  I  Q  H  Q  R  I  H  T  G  E  R

905   ccctacgtgtgtcccttgtgtgggaaagccttcaaccacagcaccgtcctgcggagccaccagagggtgcacactggggagaag
285    P  Y  V  C  P  L  C  G  K  A  F  N  H  S  T  V  L  R  S  H  Q  R  V  H  T  G  E  K

989   cctcacgggtgcagcgagtgtgggaagaccttcagcgtgaagaggacgctgctgcagcaccagcgggtgcacaccggggagaag
313    P  H  G  C  S  E  C  G  K  T  F  S  V  K  R  T  L  L  Q  H  Q  R  V  H  T  G  E  K

1073  ccctacacgtgcagcgagtgcgggaaggccttcagcgaccgctccgtgctcatccaacatcacaacgtacacactggggagaag
341    P  Y  T  C  S  E  C  G  K  A  F  S  D  R  S  V  L  I  Q  H  H  N  V  H  T  G  E  K

1157  ccgtacgagtgcagcgagtgcggcaagaccttcagccaccgctccaccctgatgaatcacgagaggattcacacgcaggagaag
369    P  Y  E  C  S  E  C  G  K  T  F  S  H  R  S  T  L  M  N  H  E  R  I  H  T  Q  E  K

1241  ccctacgcgtgctacgagtgcgggaaggcctttgtccagcactcgcatctcatccagcatcagagagtccacacgggagagaaa
397    P  Y  A  C  Y  E  C  G  K  A  F  V  Q  H  S  H  L  I  Q  H  Q  R  V  H  T  G  E  K

1325  ccctacgtgtgtggcgagtgtggccatgctttcagcgcacgccggtccctgatccagcatgagcgaatccacacaggcgagaaa
425    P  Y  V  C  G  E  C  G  H  A  F  S  A  R  R  S  L  I  Q  H  E  R  I  H  T  G  E  K

1409  cccttcagtgcacagagtgtggcaaagccttcagcctgaaagcaactctgatcgtgcacctgaggacccacacgggcgagaaa
453    P  F  Q  C  T  E  C  G  K  A  F  S  L  K  A  T  L  I  V  H  L  R  T  H  T  G  E  K

1493  ccctatgagtgcaatagctgcggcaaggcgttcagccagtactcggtgctcatccaacaccagaggatccacaccggagagaaa
481    P  Y  E  C  N  S  C  G  K  A  F  S  Q  Y  S  V  L  I  Q  H  Q  R  I  H  T  G  E  K

1577  ccctatgagtgtggcgagtgcggacgggccttcaaccagcacggccacctgatccaacatcagaaagtacacaagaagctgtga
509    P  Y  E  C  G  E  C  G  R  A  F  N  Q  H  G  H  L  I  Q  H  Q  K  V  H  K  K  L  *

1661  CCCATGGCTGTCAGAAGCCCGCACTTAGAACATGCACAAGACCTAAACAGCCTTCCTCCAGGAGGGAGTTCTGTTCCATGTGCAG
1745  AATCGTCGTTCCCTTGGGGACTCTGTTACAGTCCCATGTCCCTCCTGGGGAGAAACATTCAGCATGTGTTAAATTTATTTTCCTC
1829  AACATCTTGGAAGTTAAACTGGAGAGTAATCCTACAACTGTAGATACTGGTAAAGACAGCCATAATTCTACCTTTGTCTATTTGA
1913  CATGAATTAGTTTATTGAAGCAACTGAAAAGCTATAAAGCTTAGGGACTATTCACATTTCAAATGAAGTTCCTTTTACACAGCTT
1997  CATGACGTTCCTAAATGTATCTGATCATTCTAATATTGTGCTTAAATTTAGAATGTTGGACCATTTTCGGTATCCTTTTAAAACA
2081  CTTAAGTGCTATGTTGATAAAGTAAT***ATTAAA***AAAAAAAAAA
```

Figure 3.8 Sequence of KRAB-ZFP 492

A) Cartoon of protein domains present in KRAB-ZFP 492. Arrowhead indicates location of gene-trap insertion. **B)** KRAB-ZFP 492 cDNA sequence with corresponding protein sequence. Bold text and asterix represents the start and termination codons respectively. The KRAB A and B domains (open and shaded boxes) are indicated and the ZF motifs are underlined. Brackets designate exon/intron boundaries with conserved splice sites. The site of gene-trap integration is indicated by an arrowhead between 449 and 450 bp. The 5' and 3' UTRs are shown in capital letters with the location of the pA signal italicised in bold in the 3' UTR. Red text indicates protein sequences used to raise antibodies. Antigenic portions of the linker region, as predicted by PIX analysis, are italicised in bold. Blue text represents potential NLS. Sequencing primers used for cloning are shown by arrows. Dashed arrows indicate primers used for genotyping (Chapter 6).

cDNA clone, I deduced that KRAB-ZFP 492 gene has an mRNA of >2110 bp (Fig. 3.8B). It encodes a predicted protein of 536 amino acids (Mw= 59.87 kDa, pI= 8.39) commencing from the first ATG codon present at nucleotide 53 to the termination (TGA) codon located at nucleotide 1638. The sequence around the proposed initiating ATG is similar to the Kozak consensus sequence (A/GNNATGG) (Kozak, 1986) and a potential pA signal (ATTAAA) is found in the 3' UTR at nucleotide position 2084, and forms the start of the polyA (pA) tail.

The entire gene is separated into the standard 4-exon structure observed with most other KRAB-ZFPs (Shannon *et al.*, 1998 and references therein). The first exon contains 5' UTR, a small N-terminal sequence and the initiation codon. The second and third exons contain the KRAB A and B box respectively, whilst the last and largest exon contains the linker region, all 13 ZFs and the translation stop codon. Each of the zinc fingers conform exactly to the consensus sequence CX2CX3FX5LX2HX3H and are connected by the H/C link consensus sequence TG/QEKPY, suggestive of a role in DNA binding (Nagaoka *et al.*, 2001). This consensus is typically found in members of the KRAB C2H2-type zinc finger family. A linker region of 96 amino acids separates the KRAB domain from the ZFs and a NLS is present in the second ZF at amino acid position 219 (KHRR), although the gene-trap fusion which lacks the ZF is nuclear so either alternative NLSs exist or it is nuclear via interaction with other proteins, e.g. KAP-1. The length of the introns between each of these exons is 203 bp, 228 bp and 5053 bp respectively. I confirmed the predicted exon-intron boundaries by aligning the full-length IMAGE clone cDNA with the genomic DNA sequence (AY187287) (Fig. 3.8B). Each boundary obeys the consensus splicing signals, where gt and ag dinucleotides are located at the 5' donor and 3' acceptor sites, respectively. However, I compared the RACE sequence to both the genomic DNA and cDNA, and discovered that the gene-trap has not inserted into the final and longest intron of this gene as envisaged, but has inserted into the final exon (exon 4) probably by utilising a cryptic splice donor in the *En* intron after vector insertion, causing *En* sequence (Fig. 3.3) to be in frame with ES492 at nucleotide position 499 (Fig. 3.8). The insertion of this gene-trap

construct into the exons of trapped genes is not a novel occurrence and ~15% of clones have previously been trapped in this way (Bill Skarnes, Pers. Comm.).

3.6 Discussion

The abundance of KRAB-ZFPs in mammalian genomes may suggest why seven of such genes have been isolated of the hundred or so genes trapped in ES cells (Sutherland *et al.*, 2001). Furthermore, although many have been postulated to have a lineage-specific role during development, these genes must be expressed in ES cells to allow them to be detected by our screening system. The number and quality of KRAB-ZFP gene sequences contained in biological databases has improved over the time course of this PhD, enabling characterisation and comparison of these gene-trap proteins. The chromosomal locations of some of the gene-traps have been predicted by Ensembl based on BLAST hits with assembled contigs (section 3.3), but FISH mapping of gene-trap integration sites has proved to be vital in enforcing these Ensembl predictions. For example, Ensembl suggests that similar sequences matching the F9/30A6 gene are situated on ~10 different chromosomes, none of which are MMU9 (Fig. 3.3). Although FISH analysis has proven that the F9/30A6 gene-trap is situated on MMU9 near the telomere, this cell line was found to be abnormal and so the position of this gene may be the result of a translocation, although the FISH signal did appear to be in the same region as the MMU9 paint (data not shown). The complexity in assembling repetitive or highly similar DNA sequences in the genome is exemplified when attempting to identify the genomic location of highly similar KRAB-ZFPs, and numerous methods of gene identification will need to be employed to affirm the correct genes.

A large body of evidence suggests that KRAB-ZFPs can be arranged in clusters of tandem gene arrays that are scattered throughout many chromosomes (Bellefroid *et al.*, 1991; Tunnacliffe *et al.*, 1993; Tommerup and Vissing, 1995). These arrays of head-to-tail KRAB-ZFP genes may be the result of tandem duplications of individual genes, multiple genes and possibly whole clusters (Bellefroid *et al.*, 1991; Lichter *et al.*, 1992; Tunnacliffe *et al.*, 1993; Tommerup and Vissing, 1995). These clusters are also composed of inverted KRAB-ZFP genes and to a much lesser extent,

pseudogenes (Villa *et al.*, 1996; Krebs *et al.*, 2003). A fewer number of pseudogenes have accumulated during the expansion of KRAB-ZFP families in stark contrast to the expansion of the olfactory and MHC gene clusters. Also studies of KRAB-ZFP genes residing in many clusters have indicated that the bulk of these genes are expressed and contain a significant number of open reading frames (ORFs) (Bellefroid *et al.,* 1993; Dehal *et al.*, 2001; Looman *et al.,* 2002; Shannon *et al.*, 2003). Therefore, tandem clusters of KRAB-ZFP genes are postulated to be undergoing unusual positive selective pressures that actively favour the maintenance of gene diversity in the target recognition of duplicated copies. This suggests a role for KRAB-ZFPs in speciation (Shannon *et al.*, 2003). There are also some KRAB-ZFP genes that lie out with gene clusters in more isolated regions of the genome that are close to only a few other ZFP or KRAB-ZFP gene loci (Urrutia, 2003; ES492, this thesis). Furthermore, human KRAB-ZFPs have a biased distribution to telomeric, centromeric and/or fragile sites within the human genome (Lichter *et al.*, 1992; Eichler *et al.*, 1998). More recently, a canine homologue of human KRAB-ZFP ZNF331 (a putative target gene in thyroid tumourigenesis) has been mapped to chromosome 1q33 (CFA1q33), and cytogenic hotspots associated with canine tumous have mapped to this chromosome before (Meiboom *et al.*, 2004). The KRAB-ZFPs identified in this study highlight a similar distribution over the genome as above, in that they may also be preferentially located in telomeric and centromeric regions. Moreover, at least 4 of the KRAB-ZFP genes described here (ESKN205, ES261, ES492 and ES510) seem to be found at breaks in synteny between the mouse and human genomes. The KRAB domain of ESKN205 is paralogous to that of the *Rex2* gene and whilst this gene appears to have a rat homologue, ESKN205 does not, reinforcing the idea that some murine KRAB-ZFP clusters have expanded independently of both the human and even rat genomes after their divergence (Shannon *et al.*, 2003). The fact that human and rat orthologous genes/regions could only be found for a few of the gene-trap proteins, is compatible with previous observations that different species have substantially different repertoires of ZFP genes (Shannon *et al.*, 2003).

Three of the trapped genes match to cDNAs that appear to contain KRAB-only encoding proteins (ESKN205, ES261, F9/30A6). For ES9/MZF6D (AY149175), the Ensembl predicted gene, NM_183119, whose genomic structure is predicted by EST and cDNA clones, appears to be an alternative isoform of MZF6D as it reads through the spice site located after the MZF6D KRAB box and utilises a stop codon located in what is also the intron sequence of MZF6D (AY149175) to encode a KRAB-only containing protein. For this gene it is possible that cryptic splice sites are used to produce alternative isoforms of the same protein that either do or do not possess ZFs and therefore alter the function of the protein accordingly. This has recently been reported for ZFP208 whose KRAB-only isoform interacts with SRY (Oh *et al.*, 2005). Taken together, this finding sheds new light on the potential number of KRAB-only proteins in the genome. Also, the KRAB domain has nuclear targeting capability since all gene-trapped proteins localise to the nucleus (Sutherland *et al.*, 2001; Chapter 4) even though the predicted NLSs, which are always found in the linker/ZF domain, are removed. This is of functional relevance suggesting that it is the protein partners of the KRAB domain that recruit the gene-trapped proteins to the nucleus.

For the further study of KRAB-ZFP function I chose to clone KRAB-ZFP 492 because at the beginning of my PhD this was the only gene-trap line to which a correct gene could be assigned, and for which the surrounding region is well described. The site of insertion of the gene-trap in ES492 indicates that the ZFs of this protein are missing in the gene-trap line and is likely to represent a partial knockout of gene function in the homozygous state. Since all gene-traps potentially remove the ZF domains of the trapped proteins, it is unlikely that these mutant proteins will be able to bind their target genes. Only one transgenic KRAB-ZFP mouse line (*nrif*$^{-/-}$) had been described in the literature before the onset of my PhD (Casademunt *et al.*, 1999). I believed that some of the gene-trap lines characterised in this chapter could be used for the generation of mouse KRAB-ZFP knockouts and that these would be predicted to result in null, or even dominant negative phenotypes in the mouse since they lack ZF domains. KRAB-ZFP 492 is an A+B family member in a non-clustered part of the genome. As it does not have any close

homologues I decided to choose this KRAB-ZFP for further study. I also chose KRAB-ZFP 113 (A+C) for further studies because of its genomic location in a ZFP cluster and because ES113 was the only other novel gene at the start of my PhD that showed a good match to a cDNA clone and EST. However, I could not be certain that this match was correct due to the multiple homologues and poor sequence information at the genomic locus.

Chapter 4

Sub-cellular localisation of KRAB-ZFPs and their co-repressor, KAP-1

4.1 Introduction

At the start of my PhD, the literature described the localisation of many epitope-tagged KRAB-ZFPs to the cytoplasm, the nucleus, and sub-structures within the nucleus (De Lucchini *et al.*, 1991; Grondin *et al.*, 1997; Payen *et al.*, 1998; Huang *et al.*, 1999; Yano *et al.*, 2000; Côté *et al.*, 2001). Localisation to pericentromeric heterochromatin had also been reported in some nuclei, with this localisation being variegated within the cell population (Payen *et al.*, 1998; Sutherland *et al.*, 2001; Matsuda *et al.*, 2001). In ~80% of NIH 3T3 cells, over-expressed, epitope-tagged KRAZ1 and KRAZ2 have a nuclear diffuse staining pattern, whilst in the remaining 20%, they co-localise at pericentromeric heterochromatin with the KRAB-ZFP co-repressor, KAP-1 (Matsuda *et al.*, 2001). KAP-1 was already known to co-localise with heterochromatic and euchromatic HP1 proteins in nuclei of NIH 3T3, F9 embryonic carcinoma (EC) and P15 EC cells (Ryan *et al.*, 1999; Nielsen *et al.*, 1999). Based on an experiment that shows that Trichostatin-A (TSA), a specific inhibitor of histone deacetylases (HDACs), can partially relieve HP1 dependent silencing activity by KAP-1 (Nielsen *et al.*, 1999), Matsuda *et al.* showed that TSA treatment disperses KRAZ1 and KAP-1 away from centromeric heterochromatin (Matsuda *et al.*, 2001). In contrast, HP1α remained at centromeric heterochromatin, indicating that KRAZ1 and KAP-1, but not HP1α, are dependent on hypoacetylated histones for their targeting to heterochromatin and that this chromatin modification may precede their translocation.

Despite this data, no research had yet been published on how the association of KAP-1 and/or KRAB-ZFPs with centromeric heterochromatin could be induced and how this may result in heterochromatin-mediated gene silencing, as suggested by Ryan *et al.* in 1999 and again by Schultz *et al.* in 2001. Hence, I decided to broach this question by examining how the different stages of the cell cycle and the effects of cell differentiation affect the sub-nuclear localisation of gene-trapped KRAB-ZFPs (Sutherland *et al.*, 2001) and KAP-1. I also wanted to know if the occurrence of

KAP-1 at centromeric heterochromatin was restricted to particular cell types as inferred by ZFP-37 (a KRAB-ZFP with a truncated KRAB A domain), because antibodies to ZFP-37 localise to constitutive heterochromatin adjacent to nucleoli in occulomotor neurones (Payen *et al.*, 1998). Furthermore, during spermatogenesis, KAP-1 is preferentially associated with heterochromatic structures of specialised Sertoli cells and round spermatids, as well as with meiotic chromosomes (Weber *et al.*, 2002).

Apart from the ZFP-37 study, all localisation studies of KRAB-ZFPs utilised expression vectors in which KRAB-ZFPs were tagged to either the influenza HA (haemagglutinin) epitope (Grondin *et al.*, 1996), GFP (Yano *et al.*, 2000; Mark *et al.*, 2001), myc (Matsuda *et al.*, 2001) or other such markers. Subsequent transfections allowed visualisation of their whereabouts within the cell, but in these experiments, KRAB-ZFP expression is not under the control of the endogenous promoters so the proteins are potentially over-expressed and may mis-localise. For these reasons, I used immunofluorescence to ascertain the sub-nuclear localisations of the 7 KRAB-ZFPs trapped in the gene-trap screen (Sutherland *et al.*, 2001), since they are under the control of their endogenous promoters and are being expressed at physiological levels in the cell and in the correct temporal pattern.

The gene-trapped proteins are not full length KRAB-ZFPs (Chapter 3) and some gene-trap proteins have been found to mis-localise because they are missing portions of the endogenous protein (Sutherland *et al.*, 2004). To investigate if either the gene-trapped 492 KRAB-ZFP or GFP (enhanced green fluorescent protein)-tagged 492 constructs that I made behave in the same way as their endogenous counterpart, I raised antibodies to the endogenous protein for use in localisation studies.

4.2 Differentiation of ES/F9 cells with retinoic acid

In the majority of the KRAB-ZFP gene-trap cell lines, the β-gal-fusion proteins were found to be nuclear diffuse. However, some lines showed a heterochromatic anti-β-

gal (α-β-gal) localisation of fusion protein in a subset of cells (Sutherland *et al.*, 2001). These cells were often large, lying out-with the cell clusters and their morphology suggested that they might be differentiated.

The gene-trap screen was done in mouse E14 ES cells and F9 EC cells, that are totipotent and pluripotent respectively, and therefore undifferentiated. Gene-trapped ES cells were routinely cultured in the presence of LIF, but to confirm the differentiation status, I performed alkaline phosphatase staining (section 2.8.3.1) on two of the cell lines, ES113 and ES492 (Fig 4.1). An intense pink colour indicative of alkaline phosphatase activity is seen in undifferentiated ES cells (Matsui *et al.*, 1992). Using this assay I found that the majority of cells in the ES113 and ES492 cell lines, which were grown in the presence of LIF, formed densely packed clusters of cells and were undifferentiated (Fig. 4.1, arrowheads). However ES cells at the edge of undifferentiated cell clusters often lacked stain, and were more irregularly shaped, probably indicative of spontaneously differentiated ES cells (Fig. 4.1A, arrows).

To differentiate ES and F9 cells, I supplemented LIF free culture media with 5x 10^{-9} M retinoic acid (RA) for four or two days respectively (section 2.7.2). RA can induce differentiation in F9 EC cells (Strickland and Mahdavi, 1978) where morphological changes are first observed. Upon addition of RA, densely packed cells became much less compact and moved apart from each other as reported. Cell morphology also alters so that cells become larger, flatter and more irregularly shaped. In differentiated ES cells, alkaline phosphatase activity is diminished (Chiquoine, 1954; Donovan *et al.*, 1986) and absence of pink staining confirms that most cells have differentiated (Fig. 4.1, arrows). It is clear that undifferentiated cells remain in culture after four days of RA treatment. Even after longer periods of RA treatment, I could not achieve complete removal of these undifferentiated clusters (data not shown).

I wanted to be able to distinguish between undifferentiated and differentiated cells on a cell-by-cell basis so I performed immunofluorescence on pFa fixed ES and F9 cells

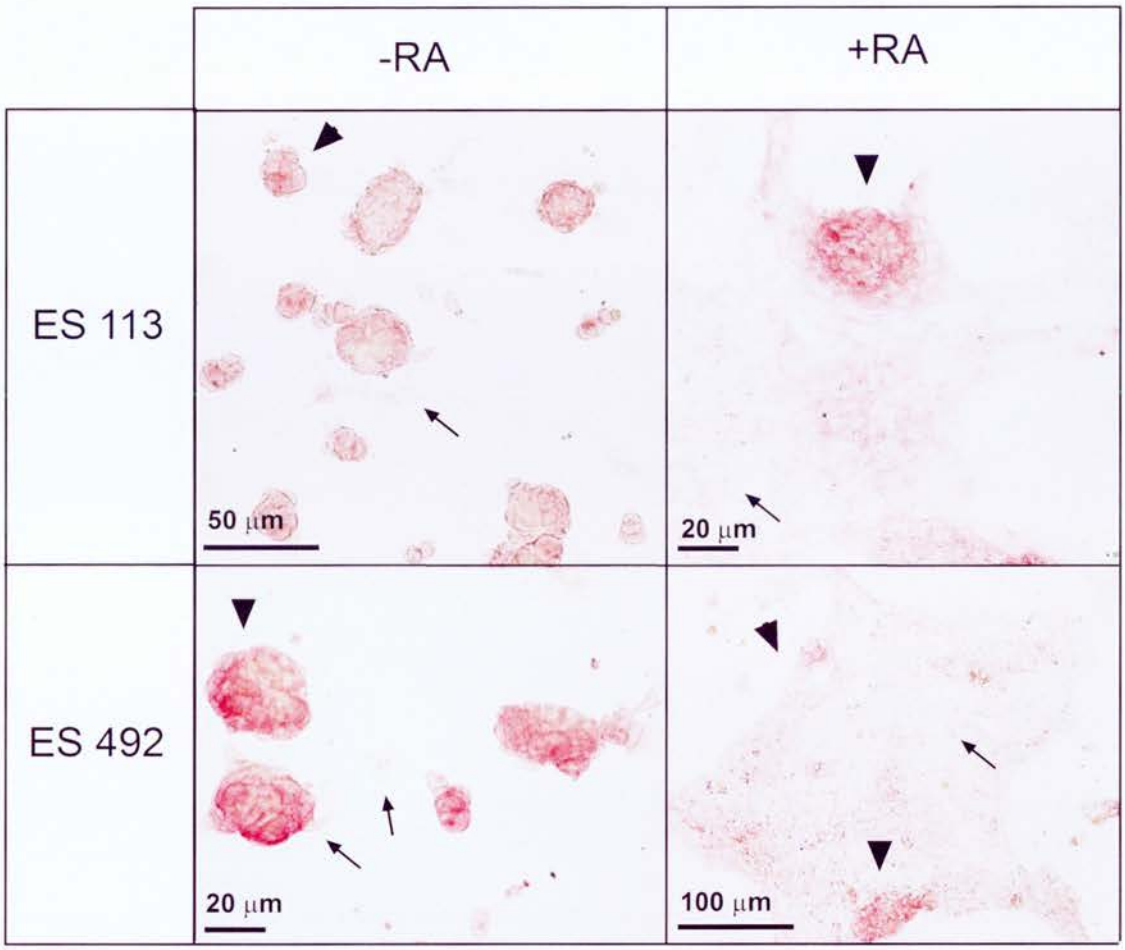


Figure 4.1 Alkaline phosphatase staining of ES cells

Before differentiation (-RA), the majorities of ES113 and ES492 cells are found in compact clusters and stain an intense pink colour indicative of alkaline phosphatase activity (arrowheads). These cells are undifferentiated. When cells are treated with RA (+RA) the majority of cells elongate, protrude extensions and lack alkaline phosphatase activity (pink stain) indicative of differentiated cells (arrows). Spontaneously differentiaing cells can be seen before RA treatment (arrows), likewise some undifferentiated cell clusters remain post RA treatment (arrowheads).

with α-SSEA-1 (stage specific embryonic antigen-1) antibody. SSEA-1 is a glycolipid that is expressed on the surface of teratocarcinoma EC cells (Solter and Knowles, 1978) and ES cells of the inner cell mass (Matsui *et al.*, 1992). In contrast with undifferentiated cells, the majority of RA differentiated ES and F9 cells do not present the SSEA-1 epitope on their cell surfaces (Fig. 4.2). I therefore concluded that this method of cell RA differentiation and detection of differentiation status with α-SSEA-1 could be used to study and compare the sub-nuclear localisations of the β-gal fusion proteins and their co-repressor KAP-1 with respect to the different differentiation status of cells.

4.3 Gene-trapped KRAB-ZFPs, and KAP-1, localise both diffusely and at pericentromeric heterochromatin in undifferentiated ES cell nuclei

The sub-nuclear localisations of gene-trapped KRAB-ZFPs were determined by immunofluorescence of pFa fixed cells (section 2.8.1), with an α-β-gal antibody that recognizes the β-galactosidase portion of the gene-trapped fusion proteins. Each cell line was also co-stained with α-KAP-1 monoclonal antibody, kindly donated by Dr D. Schultz (Schultz *et al.*, 2001), which detects endogenous KAP-1 protein in the nucleus. Consequently, the sub-nuclear staining patterns of KAP-1 and each KRAB-β-gal fusion was assessed in what were mainly undifferentiated ES or F9 cell cultures.

All KRAB-β-gal fusion proteins stained diffusely in a large proportion of nuclei in the ES cell lines 9, 113, 205, 261, 492 and 510 (Fig. 4.3). However, each of the KRAB-β-gal fusion proteins were also found at pericentromeric heterochromatin in a small percentage of these cells, which manifest as bright foci of DAPI staining in mouse cells (Fig. 4.3, e.g. ES492, arrowheads). For the rest of this chapter, I will refer to the pericentromeric heterochromatin surrounding mouse centromeres as simply heterochromatin, unless otherwise stated. Likewise, α-KAP-1 displayed a

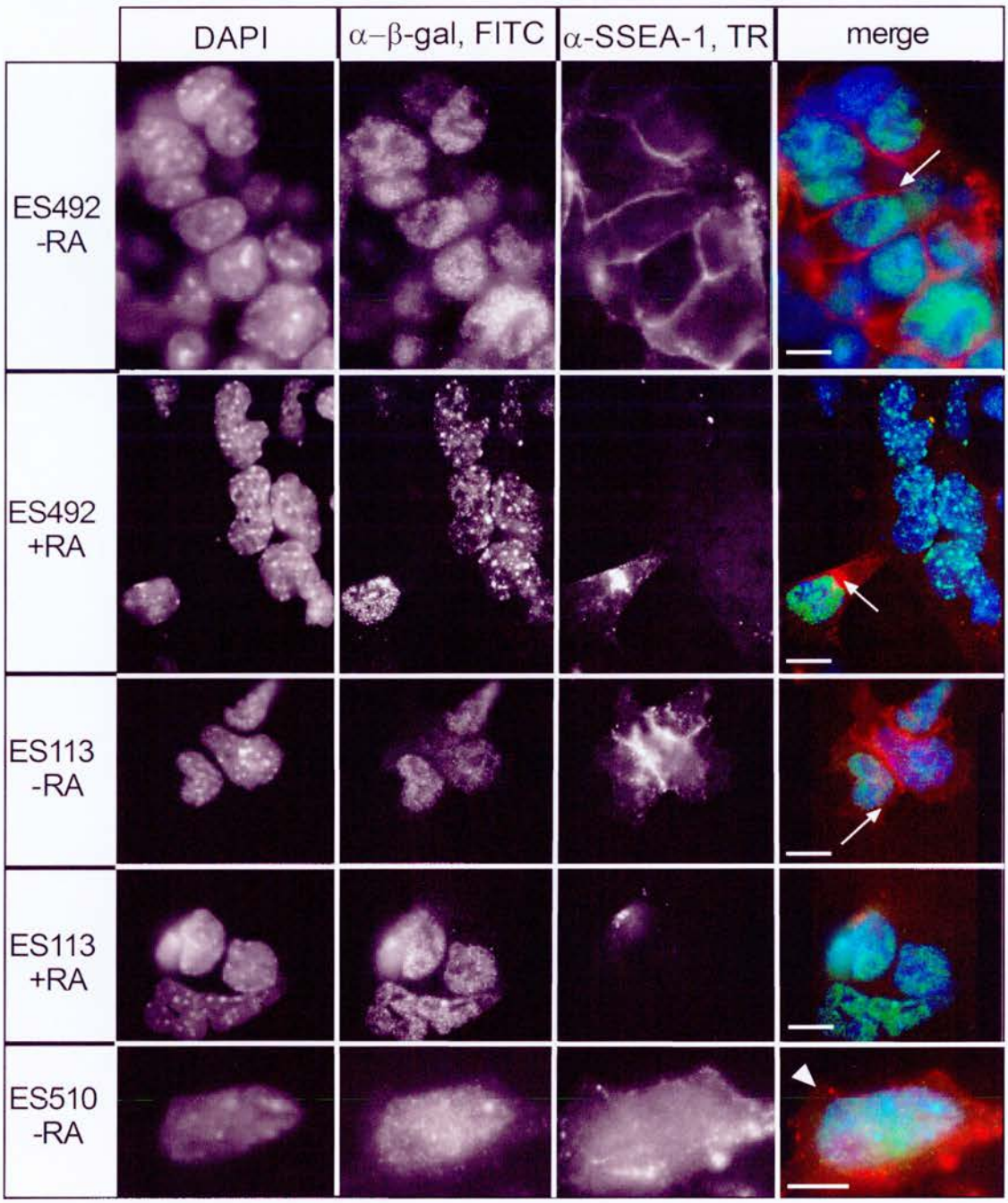


Figure 4.2 Differentiation status of individual ES cells

Immunofluorescence was performed on undifferentiated (-RA) and differentiated (+RA) cell lines with α-β-gal (green in merge) and a-SSEA-1 (red in merge) antibodies in nuclei counterstained with DAPI (blue in merge). SSEA-1 positive cells indicate undifferentiated cells, and α-β-gal stains diffusely in the majority of these nuclei (arrows). In ~1% cells, α-β-gal localises to pericentromeric heterochromatin in SSEA-1 positive cells (arrowhead). Bars = 5 mm.

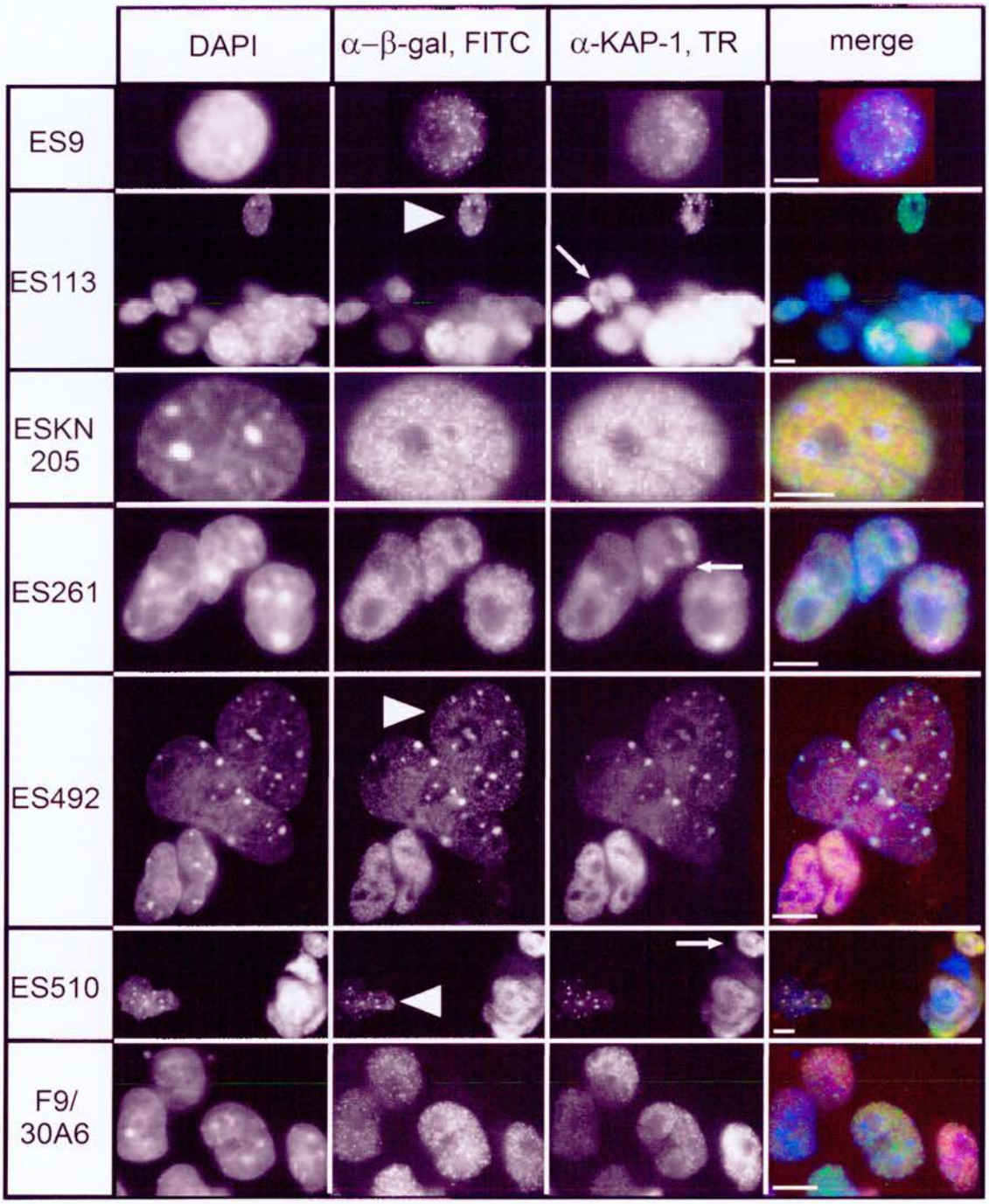


Figure 4.3 Immunofluorescence of undifferentiated gene-trapped cell lines with α-β-gal and α-KAP-1 antibodies.

Immunofluorescence of gene-trapped KRAB-ZFPs with α-β-gal (green in merge) and α-KAP-1 (red in merge) antibodies in nuclei counterstained with DAPI (blue in merge). Arrowheads indicate nuclei with co-staining of α-β-gal and α-KAP-1 at heterochromatin. Arrows indicate nuclei with heterochromatic staining of KAP-1 only. Bars = 5 mm.

similar diffuse staining pattern in most ES cell nuclei, but was also found at heterochromatin in some cells (Fig. 4.3). Moreover, KAP-1 could often be found at heterochromatin in nuclei where KRAB-ZFPs were diffusely distributed throughout the nucleus (Fig. 4.3, e.g. ES510, arrows) suggesting that KAP-1 recruitment is necessary, but not sufficient, for the localisation of KRAB-ZFPs to heterochromatin.

In the gene-trap cell line F930/A6, β-gal-fusion protein and KAP-1 have a diffuse localisation in all nuclei, and localisation to heterochromatin is never observed (Fig. 4.3).

4.4 Heterochromatic staining of KRAB-ZFPs increases upon differentiation of ES cells

To determine if co-localisation of both KAP-1 and KRAB-ZFPs at heterochromatin could be induced by differentiation I grew each ES cell line in LIF free media containing 5x 10^{-6} M RA, as previously described (section 4.1). In the F9/30A6 clone, the only KRAB-ZFP trapped in the F9 cell line, there was no increased localisation of either KAP-1 or KRAB-β-gal fusion protein to heterochromatin upon differentiation (Fig. 4.4). However, in all the ES cell lines with trapped KRAB box proteins the proportion of cells with KAP-1 and KRAB-β-gal fusion protein co-localisation at heterochromatin significantly increased upon differentiation (Fig. 4.4). As before, fusion protein was never observed at heterochromatin unless KAP-1 was present but KAP-1 was found at heterochromatin in nuclei with a diffuse distribution of KRAB-β-gal-fusion protein (Fig. 4.3 e.g. ESKN205, arrows).

I wanted to confirm that it was indeed the differentiated cells that displayed heterochromatic α-β-gal and α-KAP-1 staining, since it is known that not all cells are likely to be differentiated after RA treatment (Fig. 4.1). To identify differentiated cells, I performed co-immunofluorescence experiments with α-β-gal and α-SSEA-1 antibodies. Figure 4.2 shows that when KRAB-β-gal fusion proteins are located at heterochromatin in ES cells, SSEA-1 is not detected on the cell surface indicative of

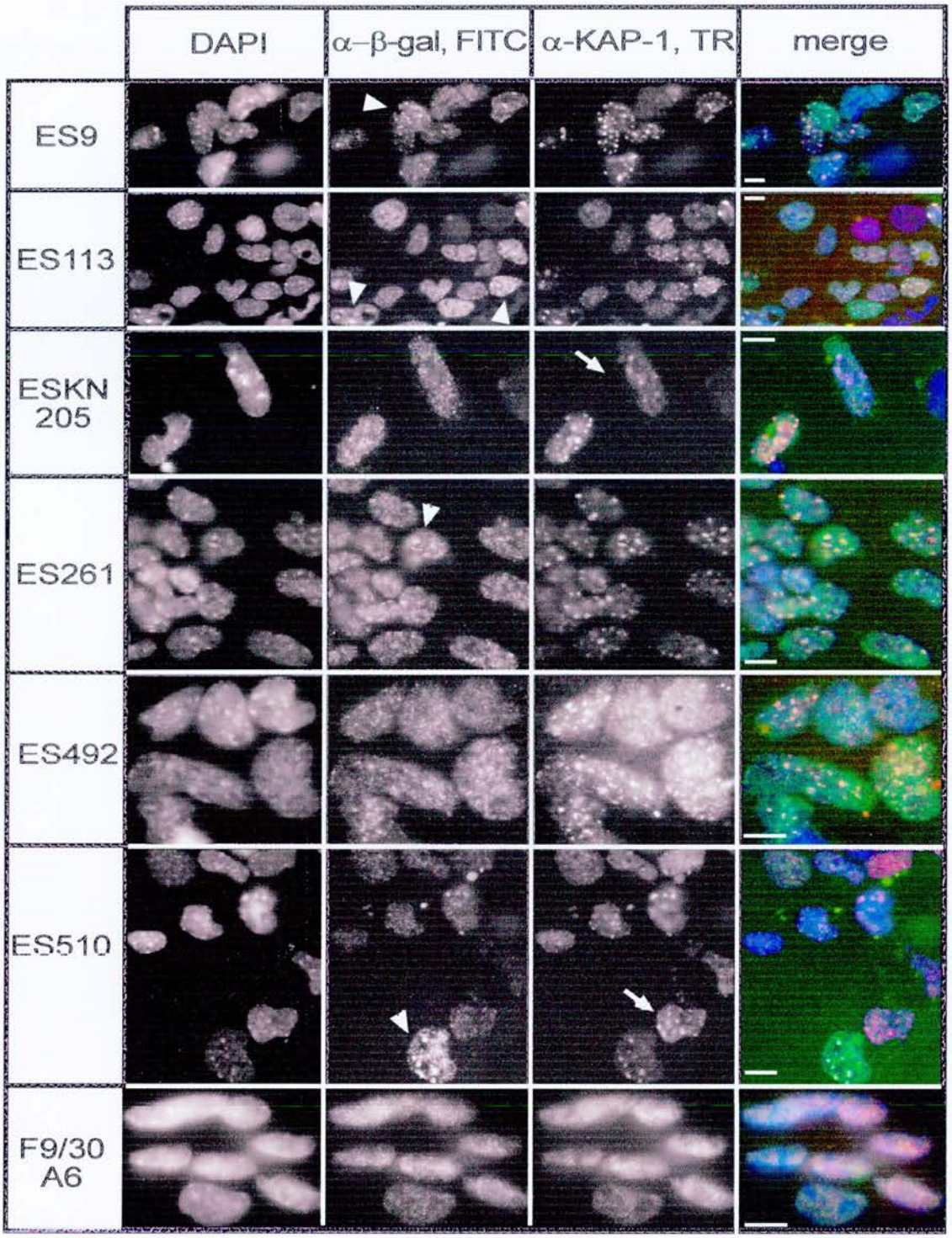


Figure 4.4 Immunofluorescence of differentiated gene-trap cell lines with α-β-gal and α-KAP-1 antibodies.

Immunofluorescence with α-β-gal (green in merge) and α–KAP-1 (red in merge) antibodies on gene-trap cells differentiated with 5x 10^{-6} M RA for either four days (ES) or for two days (F9/30A6). Arrowheads indicate α-β-gal and α-KAP-1 co-staining at heterochromatin manifested as bright DAPI foci (blue in merge). Arrows indicate nuclei with heterochromatic staining of α-KAP-1 only. Bars = 5 mm.

differentiation. Only ~1% cells expressing SSEA-1 have KRAB-β-gal fusion protein at heterochromatic foci. Dual immunofluorescence experiments were not performed with α-KAP-1 and α-SSEA-1 antibodies, because both were raised in mouse. However, previous experiments revealed that it would be highly probable for KAP-1 to be situated at heterochromatin if KRAB-β-gal fusion proteins were there (section 4.2).

In conclusion, gene-trapped KRAB-β-gal fusion proteins show increased recruitment to heterochromatin upon differentiation of ES cells with RA. In F9/30A6 cells, both KAP-1 and F9/30A6 KRAB-ZFP remain nuclear diffuse after two days of differentiation and are never found at heterochromatin. This cell line could not sustain RA treatment for four days, even when the RA concentration was lowered to $5x\ 10^{-7}$ M. SSEA-1 positive cells with KRAB-β-gal fusion protein at heterochromatin may represent transitory cells about to undergo differentiation or a population of undifferentiated cells with fusion protein, and probably KAP-1, at heterochromatin. These results support the idea that KAP-1 is required for redistribution of KRAB-β-gal fusion protein to heterochromatin upon differentiation.

Immunofluorescence did not suggest that the amount of KAP-1 or KRAB-β-gal fusion protein within the nuclei differed greatly whether the protein was found at either heterochromatic or diffuse locations.

4.5 Quantification of heterochromatic localisation

To quantify the localisation of gene-trapped KRAB-ZFPs and KAP-1, each immunofluorescence experiment on each gene-trapped cell line (undifferentiated and differentiated) was repeated three times and ~100 cells analysed each time. The number of cells with either α-β-gal or α-KAP-1 at heterochromatin was calculated as a percentage of the total number of cells tallied (Fig. 4.5). For KAP-1, percentages of cells with heterochromatic staining were counted in each ES cell line and the results pooled under the assumption that KAP-1 behaves similarly in all gene-trapped

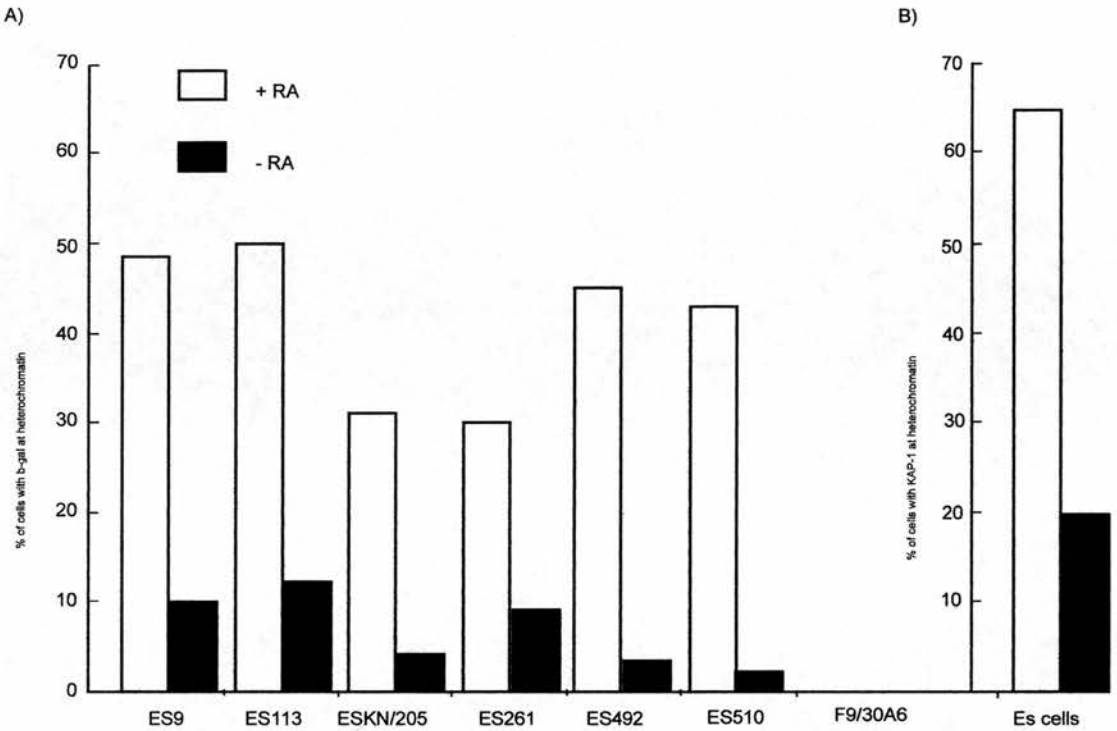


Figure 4.5 Bar chart showing the proportion of cells with heterochromatic distributions of KRAB-β-gal fusion protein and KAP-1.

A) The proportion of cells with gene-trapped KRAB-ZFPs concentrated at heterochromatin, before (-RA) and after differentiation (+RA), was calculated as a percentage of ~300 cells tallied from three independent experiments with each cell line. **B)** The average number of gene-trapped ES cells with heterochromatic KAP-1 staining from three independent experiments with each gene-trapped cell line.

cell lines. Since in some cells KAP-1 is found at heterochromatin without KRAB-β-gal fusion protein, KAP-1 is always found at heterochromatin in a higher proportion of undifferentiated and differentiated ES cells (Fig. 4.5). However, the proportion of ES cells with localisation of KAP-1 and KRAB-β-gal fusion protein to heterochromatin significantly increases with differentiation (Fig. 4.5).

4.6 Localisation of gene-trapped KRAB-ZFPs to heterochromatin is independent of the cell cycle

Increased co-localisation of KRAB-ZFPs and KAP-1 to heterochromatin upon addition of RA could result from differentiation per se, or changes in the cell cycle as rapidly dividing ES cells differentiate. When I started my PhD no studies had been published on the sub-cellular localisations of KRAB-ZFPs during different stages of the cell cycle. A link between Kid-1 and the cell cycle had been suggested, since over-expressed Kid-1, which localises to the nucleolus leads to its disintegration (Huang *et al.*, 1999). During the cell cycle nucleoli disassemble at prophase (re-appearing in telophase) and nucleolar fragmentation is also observed for other reasons, such as during apoptosis. Only three studies in the literature show KRAB-ZFPs to have potential roles in cellular proliferation. *ZF5128* (ZNF324) mRNA is up-regulated in S phase (Rue *et al.*, 2001) and GFP-tagged ZZaPK is implicated in entrance into S phase (Yang, 2002) but the sub-nuclear localisation of these proteins is unknown. ZBRK1, which binds intron 3 of *GADD45*, is likely to exert a role in the cell cycle since GADD45 is implicated in a variety of growth regulatory processes, including activation of DNA damage-induced G2/M checkpoints (Zheng *et al.*, 2000 and references therein).

To test if the gene-trapped KRAB-ZFPs localise to heterochromatin during specific stages of the cell cycle I treated both undifferentiated and differentiated cells with 0.01 M BrdU for 30 min (section 2.8.2). BrdU is a thymidine analogue that is incorporated into DNA during DNA replication, and can be detected with a specific antibody. BrdU shows either a nuclear speckle type staining or more punctate foci in

early or late S phase cells respectively. Immunofluorescence was performed on all gene-trapped cell lines and the cells stained with α-β-gal antibody. The coverslips were then removed from the slides after soaking in PBS for 30 min and a second immunofluorescence step was performed with fixation of cells in 10% formalin. These two separate fixation steps were necessary because the α-β-gal epitope was not recognised after fixation of the cells in 10% formalin. Likewise, the α-BrdU epitope is masked after only pFa fixation because it only recognises the ss DNA that is obtained with formalin fixation.

KRAB-β-gal fusion protein from ES113 and ES492 cell lines was seen at heterochromatin at both the early and late stages of S phase (Fig. 4.6, single and double arrowheads respectively) but also in non-S phase cells (Fig. 4.6, arrows). All other KRAB-β-gal fusion proteins gave the same result (data not shown). Therefore I concluded that there is no correlation between the presence of KRAB-β-gal fusion proteins at heterochromatin and S phase of the cell cycle. However, cells with heterochromatic α-β-gal staining were found in BrdU negative cells more often than BrdU positive cells (Fig. 4.6) suggesting that cells with heterochromatin-associated KRAB-β-gal fusion protein may have a decreased proliferative activity.

4.7 Localisation of GFP-tagged 492 KRAB-ZFP

Since all gene-trapped KRAB-ZFPs retain the KAP-1 interacting KRAB domain, but lack all of the zinc fingers that probably target them to specific genes or are responsible for interaction with other proteins, it is possible that they may mis-localise relative to their wild-type proteins. To determine the domains responsible for sub-cellular localisation of KRAB-ZFPs, I cloned different portions of the 492 cDNA in-frame to GFP in both the pEGFP-C1 and -N1 vectors. The primers used for PCR are described in Table 2.2.

NIH 3T3 (section 2.7.5.1) and ES cells (section 2.7.5.2) were transiently transfected with either pEGFP-C1 or -N1 vectors expressing full-length wild type (wt) 492 fused to GFP or the KRAB and ZF domains only. KRAB-ZFPs have previously been

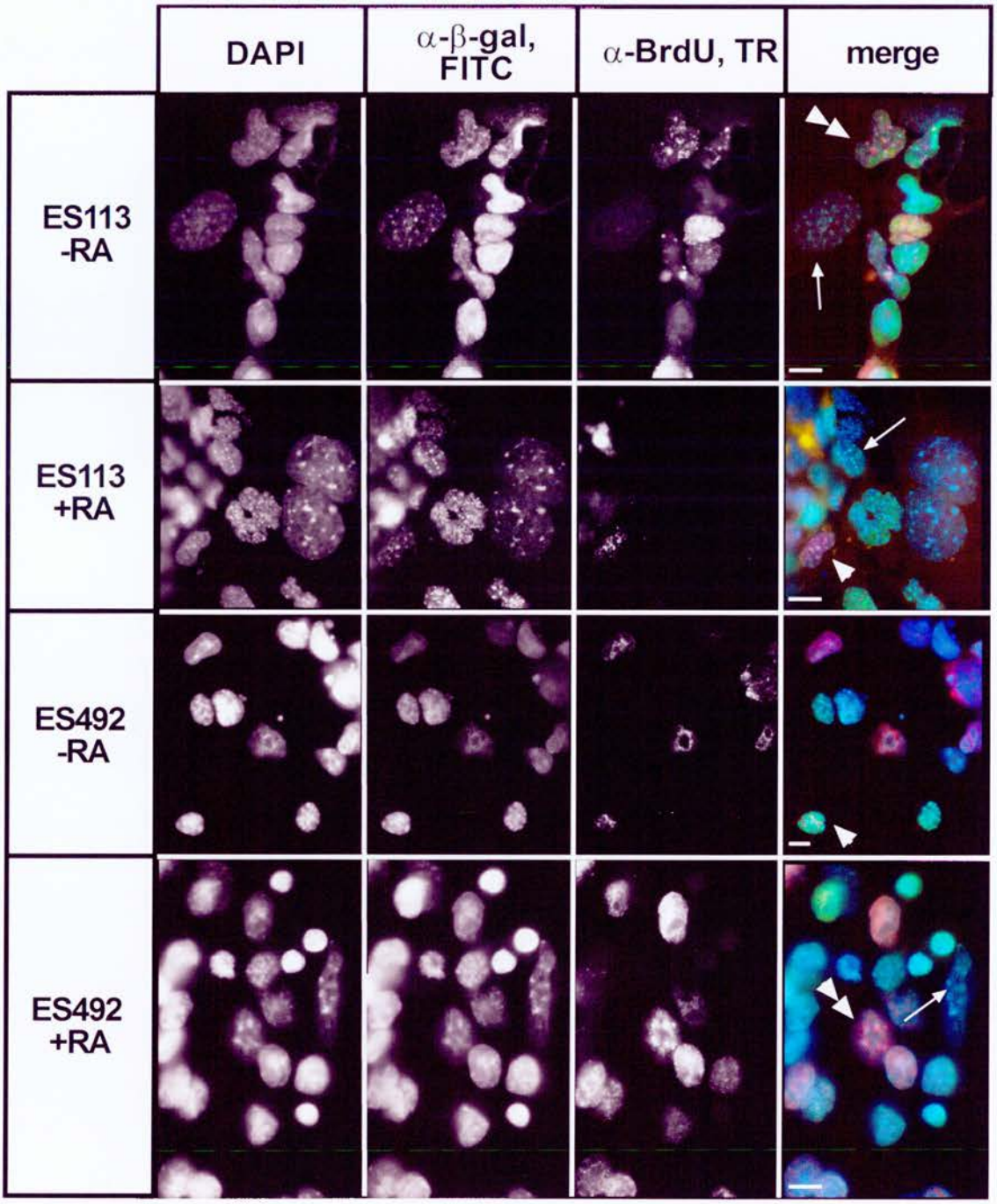
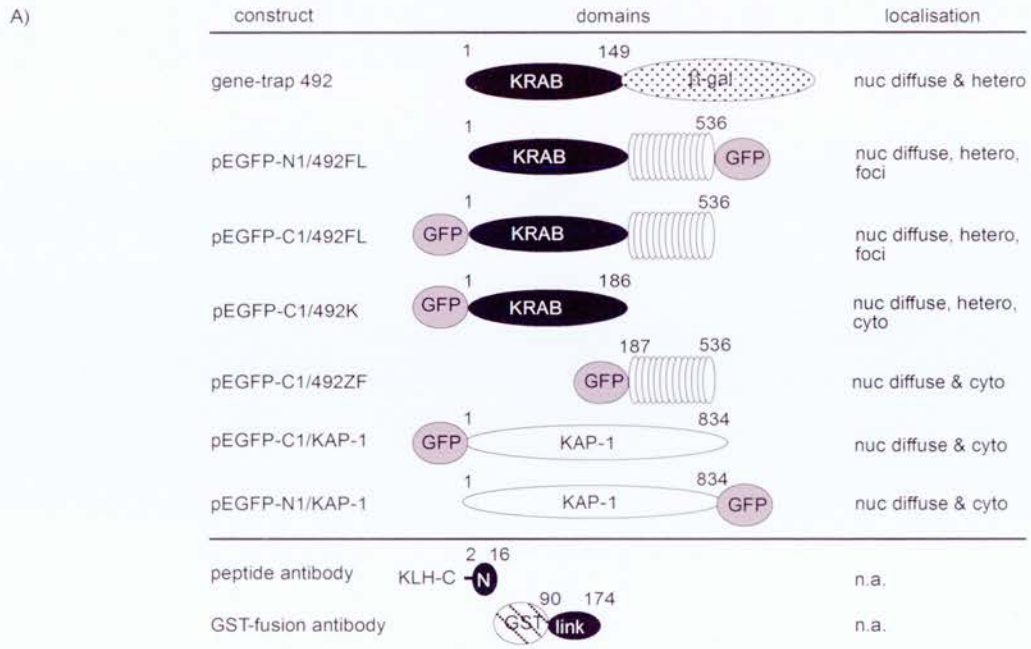


Figure 4.6 Detection of ES113 and ES492 KRAB-ZFPs in BrdU treated cells
ES113 and ES492 gene-trapped cells, before (-RA) and after (+RA) differentiation were treated with 0.01 M BrdU for 30 min and immunostained for α-β-gal (green in merged) and α-BrdU antibodies (red in merge).α-β-gal is found at heterochromatin (DAPI foci, blue in merge) in early and late S phase cells (single and double arrowheads respectively) and also in non S phase cells (arrows). Bars = 5 mm.

epitope-tagged at both their N- and C-termini (Payen *et al.*, 1998; Casademunt *et al.*, 1999; Yano *et al.*, 2000; Jheon *et al.*, 2001), I therefore decided to construct fusion proteins with GFP at both ends of the full length 492 protein to determine if the position of this tag affects its localisation (Fig. 4.7A). All constructs were fully sequenced to confirm the identity of the insert sequence.

Approximately 14-16 hours later, the cells were pFa fixed and mounted in DAPI. In cells transfected with control construct expressing GFP alone, GFP was seen throughout the cell (Fig. 4.7B). Both full-length 492 constructs (pEGFP-C1/492FL and pEGFP-N1/492FL) showed a nuclear diffuse staining pattern with large punctate spots in ~20% of cells, which co-localised with the DAPI bright spots of pericentromeric heterochromatin (Fig 4.7B, arrowheads). The staining pattern almost always excluded the nucleolus. Smaller punctate foci that did not co-localise with DAPI were often found in cells that did not contain the larger punctate foci (Fig 4.7B, arrows). The KRAB+linker construct (pEGFP-C1/492K) demonstrated a similar pattern to the full length constructs, but smaller foci and some cytoplasmic staining was also observed. This is in contrast to the KRAB+partial linker region of the gene-trap protein, suggesting amino acids 146 to 186 may be responsible for the smaller foci staining, or alternatively they may be a consequence of over-expression. The 492 ZF region alone (pEGFP-C1492ZF) showed no staining in either the heterochromatic or smaller punctate spots, but was diffuse throughout the whole cell excluding the nucleolus, indicating that amino acids 1 to 145, which comprises the KRAB domain and part of the linker, is necessary for recruitment of 492 protein to heterochromatin. The ZF region is not responsible for cytoplasmic localisation but appears to be required, in combination with the KRAB and/or linker domain, for nuclear-only distribution (Fig 4.7B).

I attempted to make stable cell lines with the constructs described but none survived the selection process. I could only obtain low transfection efficiencies with the 492 constructs and any expressing cells were usually the GFP-fusion proteins were usually expressing the GFP fusion protein at a high level (bright green cells). Many of the cells were found to be already dead or dying after fixation a day later,



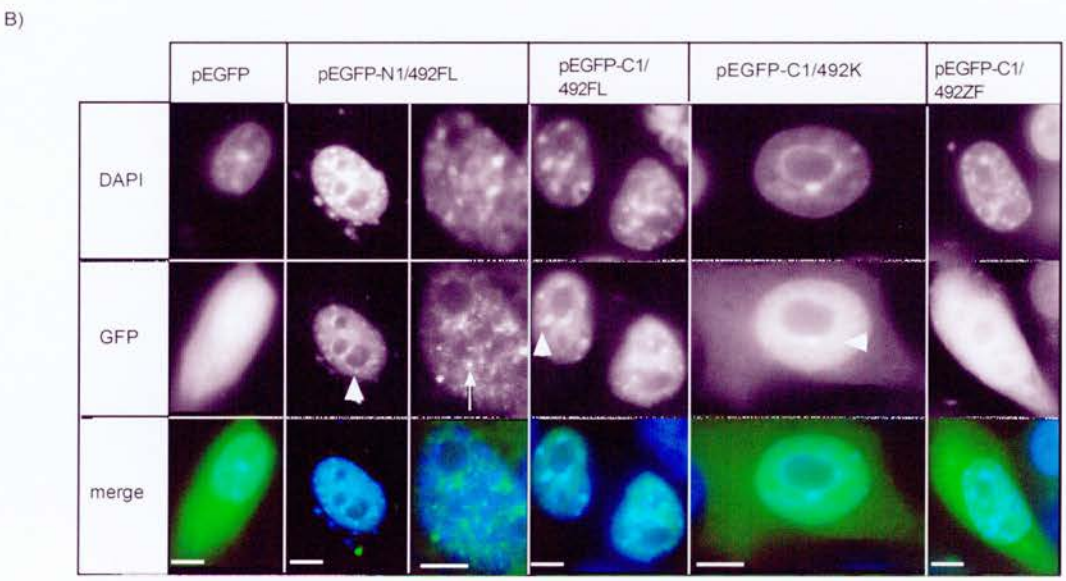


Figure 4.7 Schematic representation of GFP constructs, and localisation of their protein products in transfected cells

A) Each 492-insert sequence was cloned into either the pEGFP-C1 or –N1 vectors, which both encode GFP under control of the *lac* promoter. The name of the construct is indicated on the left and its localisation pattern on the right. *Nuc diffuse*, nuclear diffuse; *hetero*, pericentromeric heterochromatin; *cyto*, cytplasm; *foci*, small punctae foci. The KRAB-ZFP 492 N-terminal peptide sequence (N) and linker (link) region were used to raise antibodies, and are shown for comparison. Numbers represent amino acid positions in the 492 KRAB-ZFP sequence (Fig. 3.8). GFP and GST domains are not to scale. **B)** GFP-constructs were transiently transfected into NIH 3T3 or ES cells (data not shown) and their localisation patterns determined by direct immunofluorescence. Arrows indicate smaller punctate foci, arrowheads indicate large foci of pericentromeric heterochromatin. Bars = 5 mm.

suggesting that over-expression of KRAB-ZFP 492 in NIH 3T3 cells and ES cells (data not shown) is detrimental to cells. Toxicity of over-expressed KRAB-ZFPs has been previously reported for HKr18 and HKr19 (Mark *et al.*, 2001) and NRIF (Casademunt *et al.*, 1999).

I also sub-cloned KAP-1 into the pEGFP-N1 and C1 vectors to visualise its sub-cellular localisation, with a view to perhaps studying the kinetics of the KAP-1 and KRAB-ZFP 492 interaction in the cell by means of FRAP (fluorescent recovery after photo-bleaching). However, GFP-tagged KAP-1 did not localise correctly after transient transfection in NIH 3T3 cells. When GFP was tagged at either the N- or C-terminus of KAP-1 respectively, it was found to be diffuse throughout the whole cell (data not shown) and no punctate foci at DAPI bright spots were observed. Sequencing confirmed the presence of full length KAP-1 in both constructs, suggesting that mis-localisation is a consequence of either the GFP tag or over-expression. KAP-1 has previously been tagged with the HA-epitope at the N-terminus for localisation studies in COS-7 cells, which showed a homogeneous KAP-1 distribution in the nucleus (Huang *et al.*, 1999). More recently, visualisation of N-terminal FLAG tagged KAP-1in COS-7 cells was also reported to have a nuclear diffuse distribution with dotted nuclear staining (Oh *et al.*, 2005). COS-7 cells do not have visible foci of heterochromatin as marked by DAPI bright spots in human and mouse cell lines, therefore it is not known whether KAP-1 is localising to heterochromatic regions in these cells. However these tags are smaller then GFP and may be less disruptive to KAP-1 localisation. In this regard, Matsuda *et al.* did observe myc-tagged KAP-1 at heterochromatin in a sub-population of NIH 3T3 cells (2001).

4.8 Analysis of the endogenous 492 KRAB-ZFP

The localisation of GFP-tagged full length 492 is similar to that of the gene-trap (i.e. heterochromatic in a proportion of cells). Since GFP-tagged protein seems to be affected by the GFP moiety and its over-expression, I wanted to analyse the endogenous 492 KRAB-ZFPs by raising an antibody to it. Peptides and fusion proteins had previously been used to raise antibodies to five different KRAB-ZFPs

from four different species (De Lucchini *et al.*, 1991; Payen *et al.*, 1998; Katoh *et al.*, 1998; Jheon *et al.*, 2001; Tanaka *et al.*, 2002). However, only the α-mouse ZFP-37 antibody had been used to study KRAB-ZFP sub-nuclear localisation.

In the adult brain, ZFP-37 is neuron specific and associates with constitutive heterochromatin adjacent to nucleoli and/or in the interior of the nucleolus depending on the neuronal cell type (Payen *et al.*, 1998). Moreover, apart from the DNA binding ZF motif, ZFP-37 preferentially binds ds DNA via its histone H1-like DNA binding motif, which more closely resembles the carboxy termini of histone $H1^0$ and H5. These histone variants replace histone H1 in terminally differentiated cell types, and since other nucleolar components also have histone H1-like motifs (nucleolin), this suggests ZFP-37 to be included in the group of histone H1 variants. Thus, it seems possible that heterochromatic localisation could be due to its linker region rather than its KRAB domain. This seems more probable in view of the fact that ZFP-37 also contains a truncated KRAB A domain, which lacks 17 amino acids previously shown to be important for *in vitro* repression activity. The question remains, therefore, whether KAP-1 is able to recruit endogenous KRAB-ZFPs to pericentromeric heterochromatin via its interaction with the KRAB domain.

4.8.1 Antibody design to the N-terminus and linker region of KRAB-ZFP 492

Two antigens from KRAB-ZFP 492 were designed in the hope of getting a good immune response from the host animal, a GST-fusion antibody and a synthetic KLH-conjugated peptide (Invitrogen/RESgen).

Amino acids 90 to 174 (Fig. 3.8B; Fig. 4.7A), corresponding to the linker region of KRAB-ZFP 492, were fused in frame to GST, expressed in *E. coli*, and purified on a GST column (section 2.5.1) before injection into host animals. Figure 4.8A (arrows) shows an SDS-PAGE gel of IPTG induced expression of 492-GST fusion protein in *E. coli* and its abundance in the soluble protein fraction (section 2.4.1). The linker region was chosen for two reasons. Firstly, previous studies indicated that the corresponding regions of other KRAB-ZFPs have successfully been used for

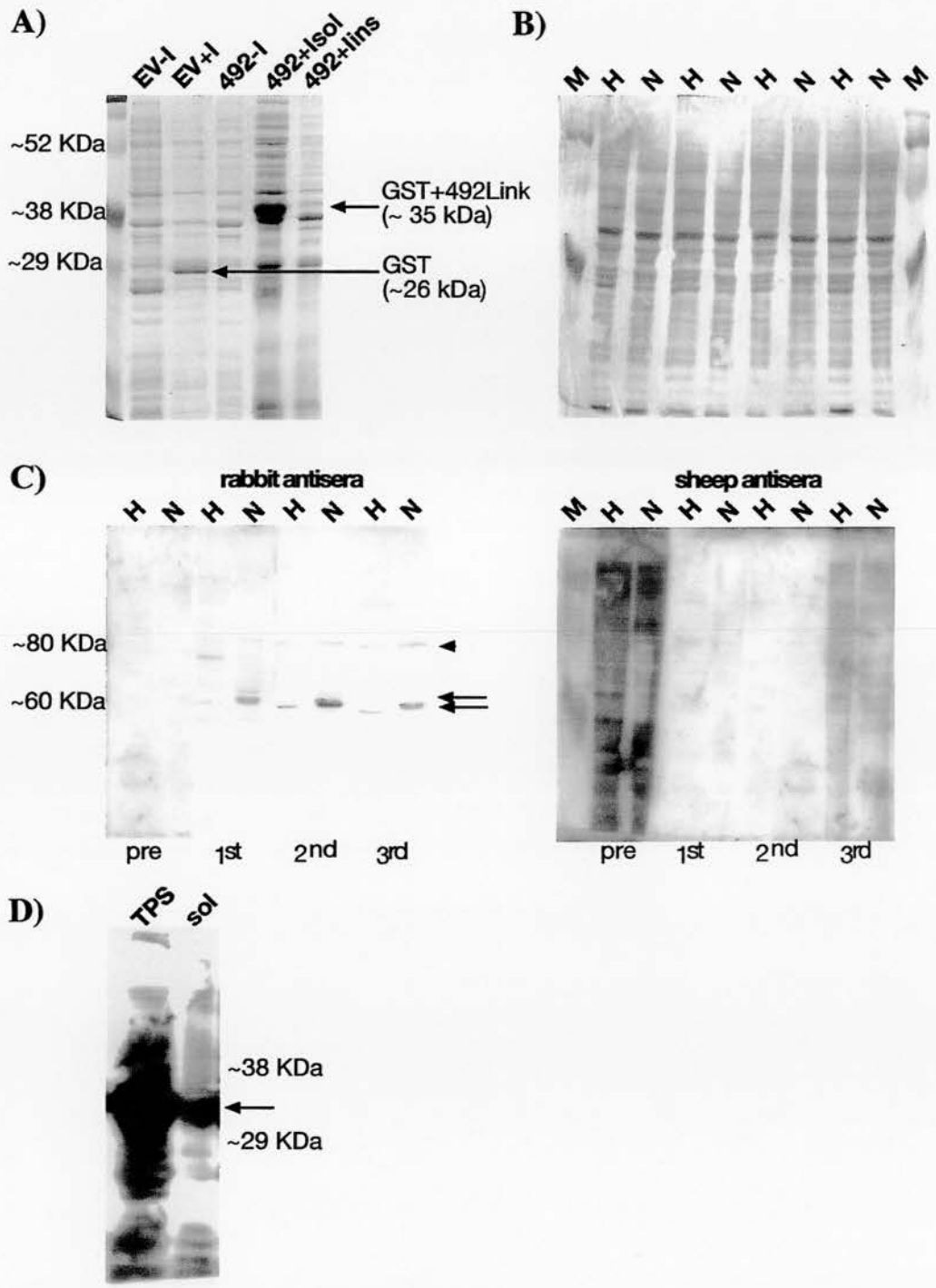


Figure 4.8 Analysis of 492 antibody production and immune response detection

A) 492-GST fusion protein expression in *E. coli* before and after 4 hours induction with IPTG (+/-I). *EV*, empty pGEX vector; *FP*, fusion protein; *sol*, soluble *E. coli* fraction; *ins*, insoluble *E. coli* fraction. **B)** An SDS-PAGE gel was stained with coomassie blue to ensure equal loading of the NIH 3T3 (N) and HeLa (H) protein extracts used in western blot analyses. **C)** Western blot analyses of H and N total protein extracts with 1:1000 dilutions of either rabbit or sheep anti-sera from their respective preimmune (pre), first (1st), second (2nd) and third (3rd) bleeds. **D)** Western blot with third bleed rabbit anti-sera (1:1000) on total protein extracts (TPS) of his-tagged 492-fusion protein (pET-492L) expressed in *E. coli*.

antibody production (Payen, *et al.*, 1998; Tanaka *et al.*, 2002) and secondly, since the size and sequence of the linker region is the most variable part of these proteins, this region is generally the most specific for a given KRAB-ZFP (Mark *et al.* 1999). A BLASTN search at Ensembl with the DNA primers used to generate the 492-linker-fusion construct (Table 2.2) show 100% identity with NM_172817 (Zfp647). Sequencing of the GST-fusion construct resulted in KRAB-ZFP 492 DNA and protein sequence with 100% identity to the corresponding region of NM_172817, whilst the next best hit (a predicted microtubule associated protein) is only 33% identical at the protein level. PIX (protein identification of unknown sequences) analyses (section 2.11) of predicted antigenic regions within the 492 KRAB-ZFP linker region highlight two antigenic regions between amino acids 108 to 116 and 137 to 144. The former of these has ~56% amino acid identity with human ZNF647 (Fig. 3.6). It was therefore possible that the antibody raised against this fusion protein could cross-react with human ZNF647.

A synthetic peptide, corresponding to N-terminal amino acids 2 to 16 of KRAB-ZFP 492 (Fig. 3.8B; Fig. 4.7A), was also chosen for antibody production because this sequence is also specific to KRAB-ZFP 492 and its human and rat homologues only. A cys residue was added at the start of the peptide sequence to allow conjugation to keyhole limpet hemocyanin (KLH), which maximises the immune response. There is 76% protein identity in this region between the mouse, rat and human homologues (Fig. 3.6). The N-terminal sequence of KRAB-ZFP AJ18 has also been successfully used for antibody production (Jheon *et al.*, 2001).

Both antigens were simultaneously injected into one sheep and one rabbit (section 2.5.2) on the premise that separation of antibodies from either host would be achieved by affinity purification if necessary.

4.8.2 Immune response detection

Serum was obtained from preimmune animals and each time subsequent to three booster injections (section 2.5.2). To determine whether a specific immune response had been elicited, western blot analyses and immunofluorescence with immunised

serum from both rabbit and sheep were performed. Total protein extracts were prepared from NIH 3T3 and HeLa cells and after SDS-PAGE resolution, proteins were transferred to PVDF membranes and the membranes incubated with serum (diluted 1:1000 with dH_2O) from each animal. The resulting western blots show that an immune response against at least one protein (perhaps two) of the predicted size of 492 (~60 kDa) was initiated in rabbit after the first booster injection and thereafter (Figure 4.8C, arrows). A second band of ~80 kDa was also detected with this α-sera compared to preimmune serum (Fig. 4.8C). Furthermore, the rabbit anti-sera detects the same protein species in both HeLa and NIH 3T3 cells suggesting the purified antibody could be cross-reactive with the human 492 homologue. Western blots using non-purified sera from the immunised sheep were dirty, so it was difficult to see whether specific antibodies had been made (Fig. 4.8C).

Finally, to confirm that the immune response elicited in rabbit was specific to the 492 protein and not GST, I made protein extracts from *E. coli* cells expressing his-tagged 492-linker TRX-fusion protein (section 4.8.3) and performed a western blot using third bleed rabbit α-sera. Only antibodies raised to the 492 portion of the 492-GST fusion protein should bind to his-tagged 492 protein at ~30 kDa (Mw. of his-492-TRX fusion protein). A strong signal of the correct size can be seen by western blotting (Fig. 4.8D, arrows), confirming the specificity of the rabbit α-sera for the KRAB-ZFP 492-linker sequence. The blot shows strong background signal because the rabbit α-serum has not been purified.

4.8.3 Purification of 492 antibodies

I decided to IgG purify and affinity purify both rabbit and sheep α-492 antibodies from final bleed sera on an IgG column as this would enable IgG antibodies raised to both the 492 N-terminus and the linker regions to be co-purified (section 2.5.3.1). 7 fractions were collected from the column during IgG purification and samples of each fraction from both animals were run on 10% SDS-PAGE gels. Heavy and light immunoglobulin (Ig) chains of ~50 kDa and ~25 kDa respectively (Fig. 4.9A) were purified from the column and the fractions pooled. The final concentrations of IgG

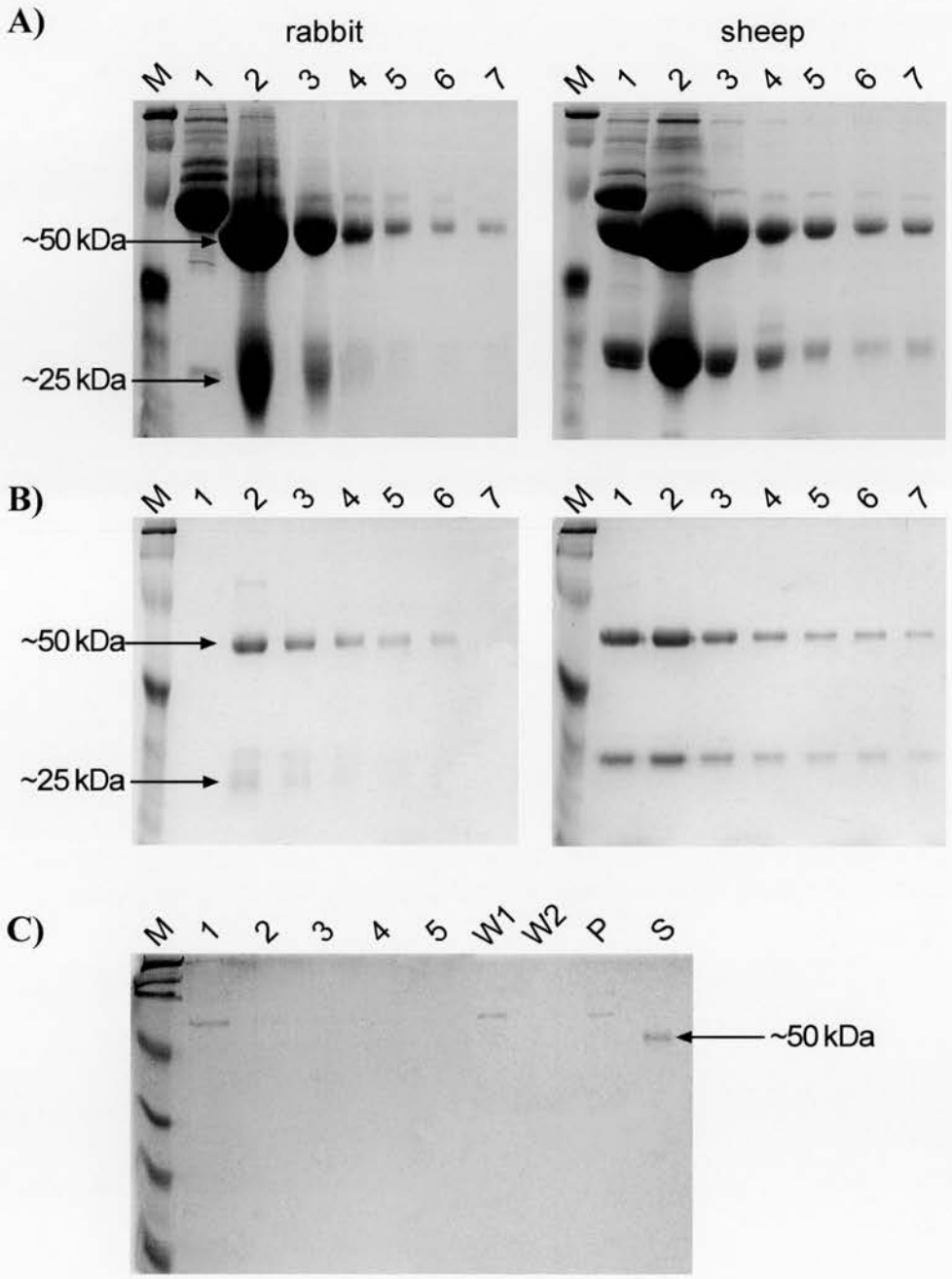


Figure 4.9 Purification of 492 antibodies

Analyses of rabbit and sheep antibodies after purification on either; **(A)** an IgG HiTrap column, for the purification of 492 IgG antibodies, **(B)** a CNBr activated column, for affinity purification of antibodies to the 492-linker region (AP492) or **(C)** a SulfoLink column, for the affinity purification of 492 peptide antibodies (AP492pep). 10 ml of each antibody fraction (numbered) was added to 10 ml of SDS-LB and boiled for 5 min. 10 ml of each protein sample was loaded on to a 10% SDS-PAGE gel. 1 mg of IgG standard was also loaded for quantification of AP492pep antibody. M, marker; W1, wash before elution; W2, wash after elution; P, 1 ml pooled fractions; S, 1 mg IgG standard.

α-rabbit and α-sheep 492 antibodies (IgG492) were measured at 40 mg/ml and 124 mg/ml respectively (section 2.5.3.1).

I also used the 492-linker region to make an affinity column for the purification of antibodies raised to the 492-linker region only (section 2.5.3.2). Firstly, the 492-linker region was expressed in E. coli as a his-tagged TRX-fusion protein (pET-492L). I then purified the fusion protein on a Ni-agarose column and then coupled it to CNBr-activated Sepharose™ 4B to produce an affinity column. Two separate affinity columns were made consisting of immobile 492-linker ionically bound to the sepharose matrix and each column was used to affinity purify either rabbit or sheep α-492-linker antibodies (section 2.5.3.3). Unbound proteins and antibodies raised to either the 492 N-terminal sequence and/or to the GST portion of the fusion protein where eluted in the flow through following washing of the column with stringent high salt buffers. 492-linker antibodies remained bound to the column until they were eluted with a low pH buffer, which altered the degree of ionisation of the charged groups at the binding sites between the adsorbent and the antibody. Protein fractions eluted from each column were pooled (Fig. 4.9B) and their final concentrations measured at 4.75 mg/ml for affinity purified rabbit antibody (AP492) and 7.27 mg/ml for the sheep antibody.

I also attempted to make an affinity column with the 492 peptide (section 2.5.3.4), by immobilising it to Sulpholink gel via its reduced cys residue. This column was used to purify α-rabbit 492 peptide antibodies (AP492pep) from the third bleed anti-sera that recognise epitopes contained within the peptide. However, purification of 492 peptide antibodies does not appear to have been successful as the main protein flushed off the column is larger than the heavy chain IgG band (Fig. 4.9C). The protein eluted from the column could be IgM antibodies or a non-specific serum protein but I did not check the identity of this band as only a small amount (~1 μg/μl by IgG standards) was eluted.

4.8.4 Characterisation of α-492 antibodies by western blot analyses

Each of the purified antibodies (IgG492 and AP492) from rabbit and sheep, and the rabbit AP492pep were tested on western blots of NIH 3T3 and HeLa nuclear and whole cell extracts. Specific bands could not be seen with either of the sheep antibodies, except perhaps for a faint band detected with a 1:100 dilution of the IgG antibody (Fig. 4.10B, asterix) suggesting that the immune response elicited by this animal was not very good. Using both the rabbit IgG492 and AP492 antibodies two specific protein bands close to the expected size of ~60 kDa were detected on NIH 3T3 and HeLa whole cell extracts, with the lower of these two bands being the most intense (Fig. 4.10A, arrows). These bands are approximately the same size bands as those detected with the third bleed serum before purification (Fig. 4.8B). A 1:1000 dilution of rabbit AP492 and a 1:1500 dilution of rabbit IgG492 were found to be optimal for western blots (Fig. 4.10B). Interestingly in nuclear extracts of both cell types, only the upper band of the ~60 kDa doublet is observed (Fig. 4.10C, arrowhead), suggesting that in addition to this larger nuclear protein, AP492 and IgG492 recognise a cytoplasmic protein. The intensity of the lower band suggests that this is the more abundant form of the protein. This band may represent a different 492 isoform as different KRAB-ZFP isoforms have been shown to exist by alternative splicing (Dreyer *et al.*, 1999; Côté *et al.*, 2001; Li *et al.*, 2003; Resch *et al.*, 2004; West *et al.*, 2004; Oh *et al.*, 2005) and although KRAB-ZFPs are usually thought to be nuclear proteins, there are examples of those with a cytoplasmic localisation (De Lucchini *et al.*, 1991; Casademunt *et al.*, 1999). Alternatively, the cytoplasmic band may be the result of post-translational processing or the nuclear isoform itself may be post-translationally modified. Finally the isoform may cross-react with a non-specific cytoplasmic protein. I have not yet distinguished between all these possibilities.

A larger protein of ~80 kDa was also detected with the rabbit IgG492 antibody in both NIH 3T3 and HeLa cell extracts. This band was also prominent in the immunised serum before purification (Fig. 4.8C) and in the nuclear extracts (Fig.

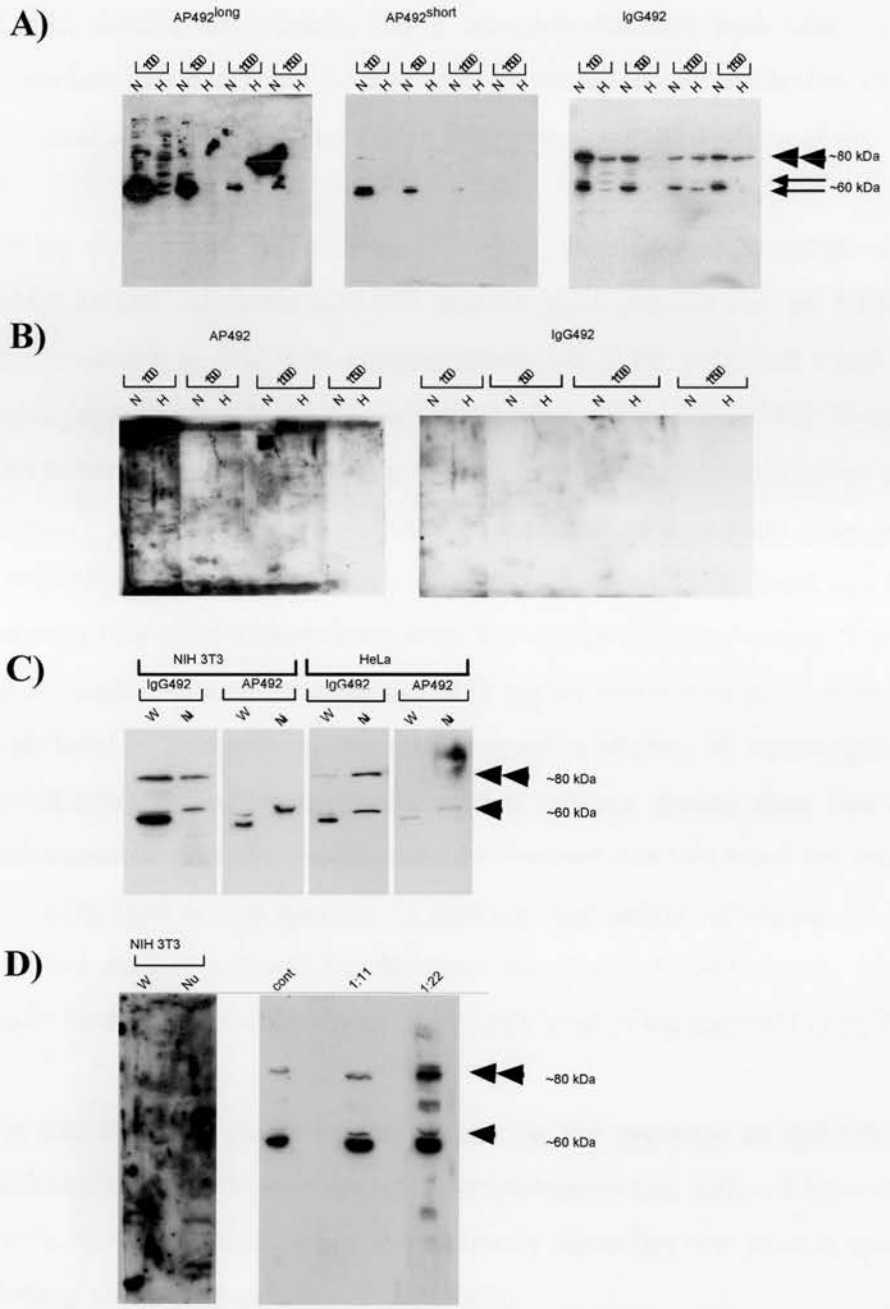


Figure 4.10 Analysis of purified 492 antibodies by western blotting
Western blots of NIH 3T3 (N) and HeLa (H) whole cell extracts with different dilutions of rabbit **(A)** and sheep **(B)** AP492 and IgG492 antibodies. In **(A)**, long and short exposures are shown for rabbit AP492 to illustrate the specificity of the antibody for the ~60 kDa bands (arrows). An asterix denotes a possible band in **(B)**. **C)** Western blots with rabbit AP492 and IgG492 antibodies on NIH 3T3 and HeLa whole cell (W) and nuclear extracts (Nu). Single arrowheads denotes the nuclear specific band. Double arrowheads indicate the ~80 kDa band. **D)** The left panel shows a western blot of rabbit AP492pep antibody on NIH 3T3 whole cell (W) and nuclear extracts (Nu). The right panel shows a competition assay between rabbit IgG492 and unconjugated 492 peptide. Western blots of total NIH 3T3 protein extracts were incubated with a 1:11 and 1:22 molar excess of IgG492:unconjugated peptide. *Cont*, control. The ~80 and ~60 kDa bands are marked as before.

4.10C, double arrowhead), but it was not detected with rabbit AP492 antibody. Therefore this band may be non-specific or due to 492 antibodies recognising the N-terminal sequence, and therefore a larger or modified isoform of 492.

To try and further characterise this band, I performed competition assays between rabbit IgG492 antibody and 492 unconjugated peptide (i.e. no KLH). IgG492 was incubated o/n at 4°C with approximately 1:5, 1:10, 1:20 and 1:100 molar excess of unconjugated peptide so that any antibodies raised to the 492 N-terminal sequence would bind the unconjugated peptide leaving only free antibodies to the 492-linker region. A western blot of NIH 3T3 whole cell extracts was probed with this competition mix and the bands analysed. The ~80 kDa band is still detected on a western blot after competition with the molar excesses tested (Fig. 4.10D, double arrowheads). Competition assays with higher molar excesses resulted in blots with a high level of background due to non-specific binding of unconjugated peptide (data not shown). I therefore concluded that either a greater than 100 molar excess of unconjugated peptide was required to compete out this band but most probably this ~80 kDa band is non-specific. I also checked rabbit AP492pep on NIH 3T3 whole cell and nuclear extracts for the presence of an ~80 kDa band. No specific bands could be detected and the blots had a high level of background (Fig. 4.10D).

For this thesis, I chose to concentrate on the presence of KRAB-ZFP 492 in the nucleus; therefore in the majority of experiments that follow I have chosen to use the rabbit AP492 antibody since this antibody identifies one protein species by western blotting on nuclear extracts (Fig. 4.10C).

4.8.5 Characterisation of endogenous 492 protein localisation by immunofluorescence

Initially I used both the rabbit IgG492 (1:1500) and AP492 (1:100) antibodies to examine the sub-cellular distribution of 492 KRAB-ZFP by indirect immunofluorescence in NIH 3T3 and HeLa cells. Cells were pFa fixed as previously described and stained with either IgG492 or AP492. AP492 and IgG492 were both found to have a grainy, nuclear diffuse distribution in both NIH 3T3 and HeLa cells,

with some weak cytoplasmic staining (Fig. 4.11). In 30-40% cells, there was strong 492 staining in the DAPI 'holes', which often correspond to nucleoli. This nucleolar staining was confirmed by co-staining with an antibody against fibrillarin, a marker of the dense fibrillar component of the nucleolus (Fig. 4.11, arrows). Although there is not complete localisation (as the nucleolus is composed of a number of sub-compartments), 492 is clearly present in the nucleoli of a subset of cells.

Previously, KAP-1 was shown to have an even, granular pattern in the nucleus of NIH 3T3 cells with localisation to pericentromeric heterochromatin in ~40-50% of cells and exclusion from the nucleolus (Ryan *et al.*, 1999; Nielsen *et al.*, 1999; Matsuda *et al.*, 2001). I also observed this result (data not shown). Punctate foci of 492-staining were also observed in ~5-10% of NIH 3T3 cells (Fig. 4.11, arrowheads). Co-staining with α-KAP-1 showed that whilst some of these foci associated/appeared adjacent with KAP-1, they did not always overlap. Interestingly, in contrast to that observed in the ES492 gene-trap cell line, neither IgG nor affinity purified 492 antibodies co-localised with KAP-1 at heterochromatin in NIH 3T3 cells. In HeLa cells, both 492 and KAP-1 showed a fine speckled pattern, but strong association with KAP-1 was not observed. This even, speckled pattern in HeLa cells has been previously recorded for KAP-1 (Underhill *et al.*, 2000).

Rabbit IgG and AP492 antibodies were characterised further by immunofluorescence on undifferentiated, pFa fixed wt ES cells and ES492 gene-trap cells. Although both antibodies show the same immunofluorescence patterns, I have only shown the results for AP492 as this gave the cleanest results. In wt E14 ES cells, AP492 stains diffusely in the nucleus and ~80% cells show a strong nucleolar staining pattern (Fig. 4.12, arrows). A small amount of cytoplasmic staining was sometimes apparent, although the nucleus occupies the majority of the volume of an undifferentiated ES cell. Surprisingly no heterochromatic staining of 492 was observed, unlike the gene-trap cell line, even after differentiation with RA. I did notice however, that some differentiated wt ES cells had small discrete foci similar to those observed in NIH 3T3 cells (data not shown). Using the ES492 gene-trap cell line, AP492 was nuclear

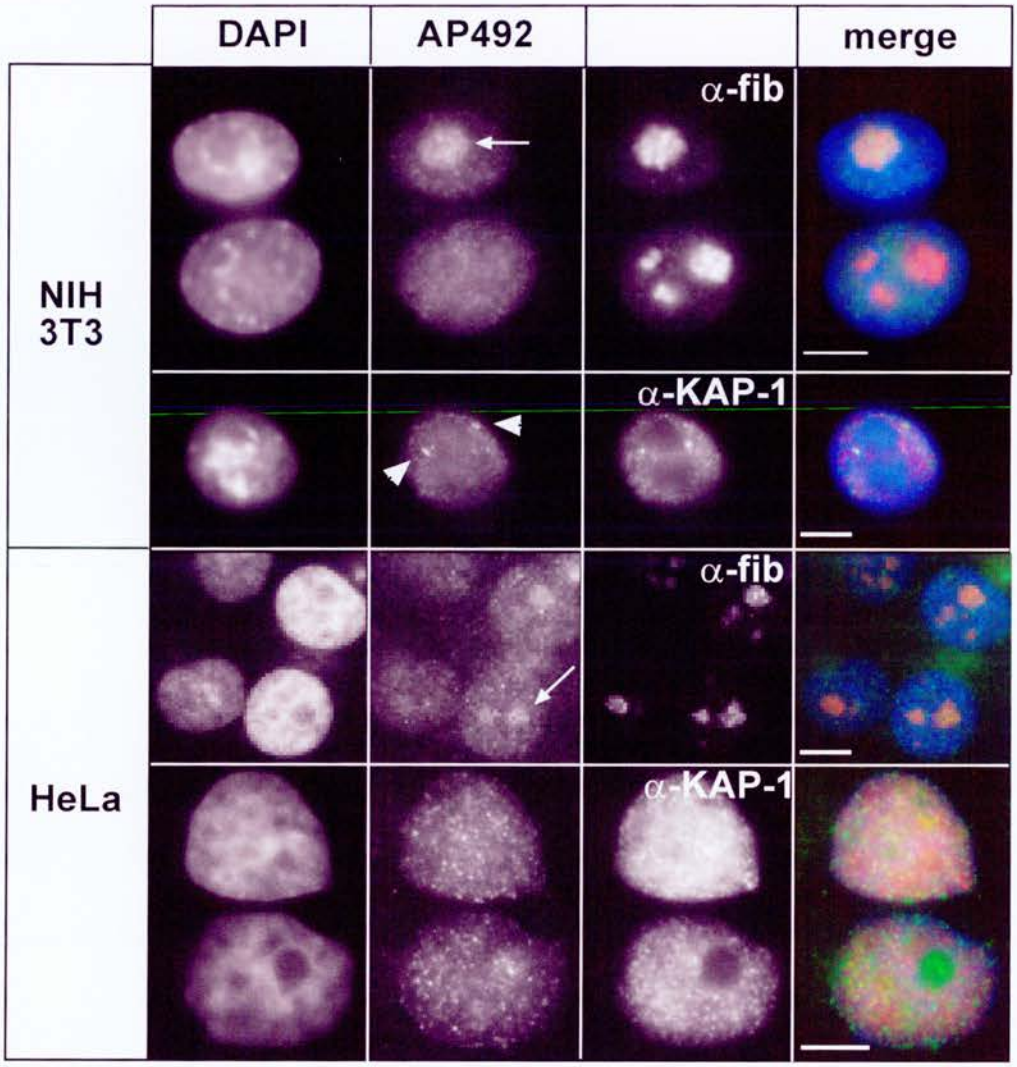


Figure 4.11 Endogeneous 492 localisation in NIH 3T3 and HeLa cells
Immunofluorescence of NIH 3T3 and HeLa cells with 1:100 AP492 (green in merge) and either α-KAP-1 or α-fibrillarin (α-fib) (red in merge) antibodies. Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining. Arrowheads indicate small punctate foci. Bars = 5 μm.

diffuse and in the nucleolus in undifferentiated cells (Fig. 4.11B, arrows). In RA differentiated cells AP492 shows a granular, nuclear diffuse distribution, with co-localisation at heterochromatin with α-KAP-1 in a similar proportion of cells as that observed previously with α-β-gal and α-KAP-1 co-staining (Fig. 4.12, double arrowheads). This suggests that the rabbit 492 antibodies recognise both endogenous 492 and the gene-trapped 492 KRAB-ZFP as a β-gal fusion protein that is recruited to heterochromatin. Since the AP492 antibody was affinity purified against the linker region, the epitope recognised by this antibody must be in the portion of the linker contained within the gene-trapped protein. Co-staining with α-β-gal and either AP492 or IgG492 antibodies was not performed since the available antibodies were both raised in rabbit. To ensure that the AP and IgG492 antibodies were specific and did not recognise any of the other gene-trapped KRAB-ZFPs, immunofluorescence was performed on ES113 cells. In undifferentiated ES113 cells, ~60-70% of cells had a strong nucleolar staining pattern with even staining in the rest of the nucleoplasm (Fig. 4.12, arrows). After differentiation of ES113 cells, immunofluorescence with AP492 showed a more diffuse staining pattern in the nucleus with strong nucleolar staining in a reduced 30-40% of cells (Fig. 4.12, arrows). If these antibodies were picking up the gene-trapped ES113 protein then the majority of these cells would be showing a heterochromatic staining pattern. These results show that the 492 antibody I have generated is specific to KRAB-ZFP 492 in the nucleus. Furthermore, these results suggest that the gene-trapped 492 KRAB-containing fusion-protein is mis-localised in the nucleus compared to the endogenous protein.

Since neither the gene-trap or GFP-tagged 492 protein was found to localise to the nucleolus, I confirmed presence of the endogenous 492 protein there by western blotting on HeLa (gift, A. Lamond) and NIH 3T3 (H. Sutherland, Pers. Comm.) nucleolar extracts with both AP492 and IgG492 antibodies. HeLa nucleolar fractions were prepared and their purity checked as described by Anderson *et al.*, 2002. NIH 3T3 fractions were prepared using the same method of fractionation and checked by western blotting with α-fibrillarin, which shows enrichment of fibrillarin in nucleolar fractions (Anderson *et al.*, 2002). In nucleolar and nucleoplasmic extracts

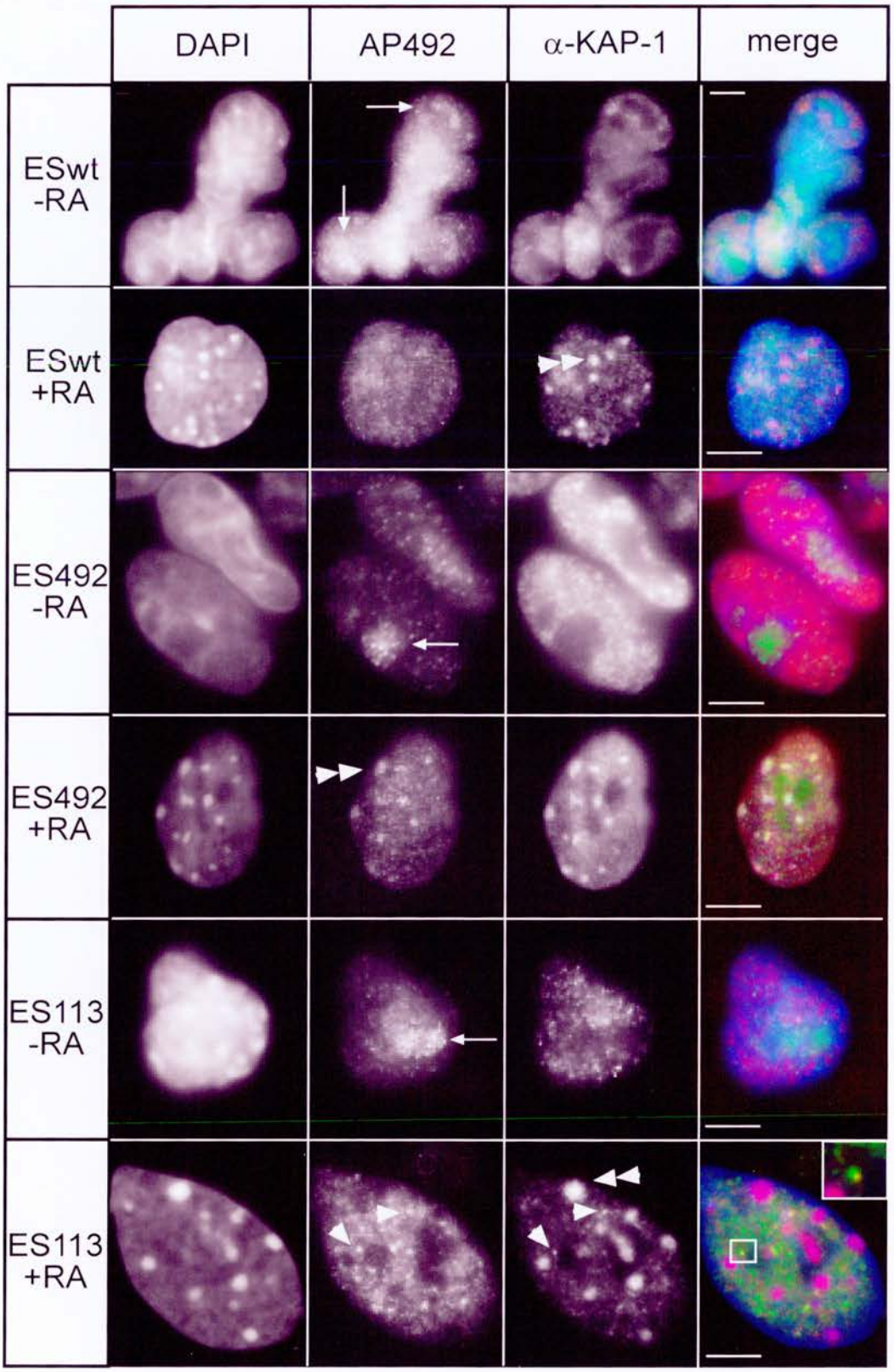


Figure 4.12 Endogeneous 492 localisation in ES cells

Immunofluorescence of wt and gene-trapped ES113 and ES492 cells, before (-RA) and after differentiation (+RA), with 1:100 AP492 (green in merge) and α-KAP-1 (red in merge) antibodies. Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining. Arrowheads indicate punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 μm.

from both cell types, AP492 detects the same ~60 kDa band as in the nuclear extracts (Fig. 4.13, arrowhead). The cytoplasmic HeLa extract does not appear to detect either of the ~60 kDa bands, however it is not known how much protein was in this extract and I only had limited amounts of material. Further characterisation of the cytoplasmic band in NIH 3T3 cells would be interesting. IgG492 detects the ~80 kDa band in all HeLa and NIH 3T3 cell extracts (Fig. 4.13, double arrowhead), as well as the ~60 kDa bands detected by AP492, suggesting again that the ~80 kDa band is probably non-specific. These results biochemically confirm that endogenous KRAB-ZFP 492 is indeed found in the nucleolus of both HeLa and NIH 3T3 cells as suggested by immunofluorescence. Hence, endogenous distribution of KRAB-ZFP 492 appears to differ from the gene-trap and GFP-tagged 492 protein.

Endogenous KRAB-ZFP 492 has a diffuse, granular pattern throughout the nucleoplasm in which small punctate foci can be found, depending on the cell type and the effects of RA differentiation. Punctate foci were also observed with 492-GFP in some NIH 3T3 cells (Fig. 3.7), which suggests that the GFP tag does not interfere with 492 localisation to the smaller foci, assuming that these are the same foci. Nevertheless, the foci appear to require the KRAB and/or linker region since they are not observed with the GFP-492-ZF construct. Thus neither the 492-β-gal fusion protein nor GFP-tagging of 492 accurately portrays its endogenous distribution. In ES cells, strong nucleolar staining is found in a higher proportion of undifferentiated cells than in differentiated ES cells or the fibroblastic NIH 3T3 and HeLa cell lines. Cytoplasmic staining is also more prominent in the latter but this could be a result of these cell types having larger cytoplasmic volume.

These results led me to investigate whether or not there is a correlation between the nucleolar localisation of KRAB-ZFP 492 and the presence of discrete KRAB-ZFP 492 foci with cell differentiation using the AP492 antibody.

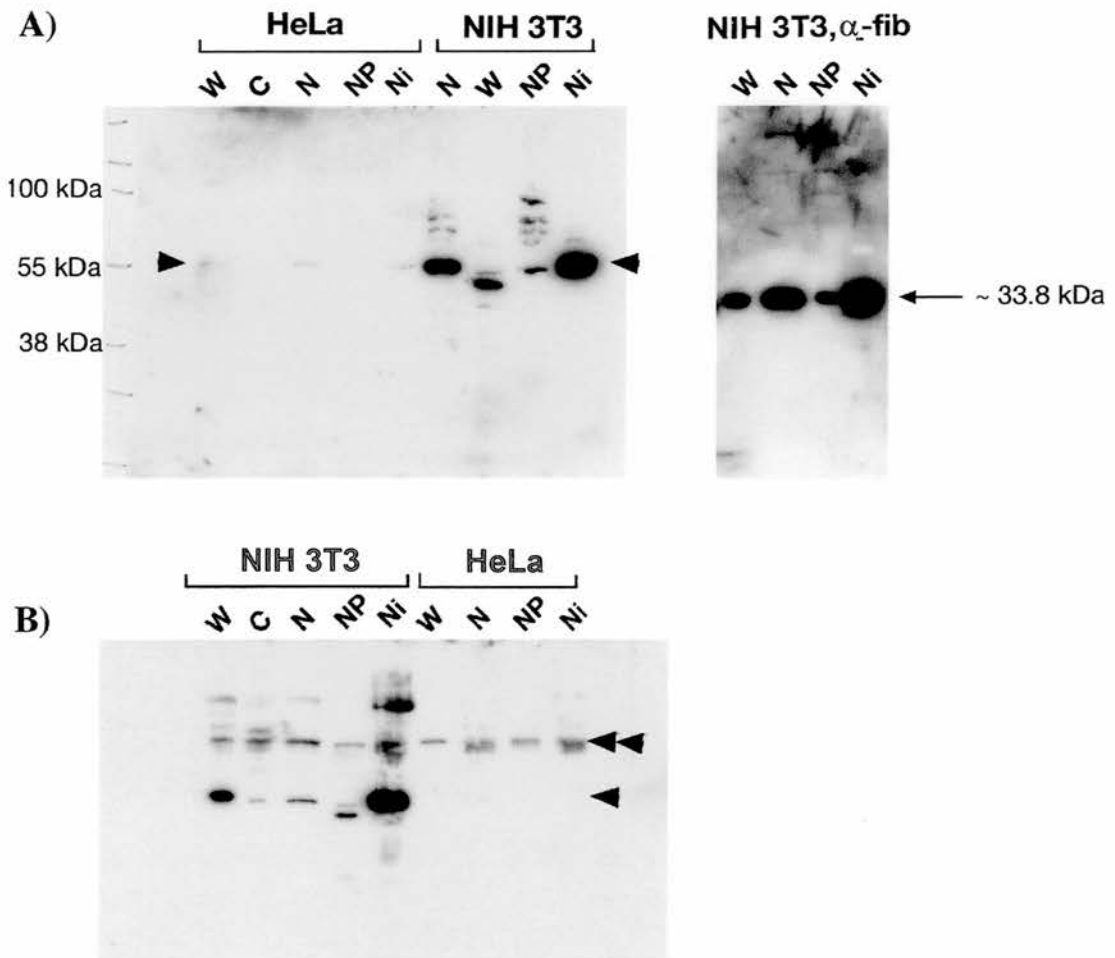


Figure 4.13 Analyses of nucleolar distribution by western blotting
Western blots of NIH 3T3 and HeLa whole cell (W), cytoplasmic (C), nuclear (Nu), nucleoplasmic (NP) and nucleolar (Ni) extracts with 1:1000 and 1:1500 dilutions of AP492 (**A**) and IgG492 (**B**) respectively. Arrowheads and double arrowheads indicate the ~ 60kDa and ~80 kDa 492 proteins. NIH 3T3 extracts were also incubated with a-fibrillarin antibody (1:5000) as a positive control (arrow).

4.8.6 Nucleolar localisation of KRAB-ZFP 492 is not dependent on differentiation

To find out if nucleolar localisation of KRAB-ZFP 492 is related to cell differentiation I used OS25 ES cells (E14 derived) (a gift of A. Smith). These cells have *β-geo* integrated into the *Sox2* locus and hygromycin thymidine kinase (*tk*) into the *Oct4* locus by homologous recombination (Billon *et al.*, 2002). Undifferentiated, (*Sox2* positive) ES cells can be selected with G418 and RA-differentiated cells selected with gancyclovir (section 2.7.4). The differentiation status of the ES cells were ensured by alkaline phosphatase staining (data not shown) before plating onto slides and fixing in 3% pFa for immunofluorescence. Undifferentiated OS25 cells were co-stained with AP492 and α-KAP-1 antibodies. Signal from AP492 is present in the nucleolus in ~80% of undifferentiated OS25 cells (Fig. 4.14, arrows). KAP-1 stains diffusely in the nucleus in ~80-90% of OS25 cells but can be found at pericentromeric heterochromatin in the remaining cells. OS25 cells were differentiated with RA (5x 10-6 M) for two days then RA in the presence of gancyclovir (2.5 μM) for a further two days. Differentiation status was confirmed by alkaline phosphatase staining (data not shown) and by immunofluorescence with α-SSEA-1 antibodies. KRAB-ZFP 492 appears to be in the nucleolus of ~26% of differentiated cells and α-KAP-1 was present at pericentromeric heterochromatin in ~80-90% cells (Fig. 4.14, double arrowheads), but KAP-1 was often found to co-localise with AP492 in smaller punctate foci (Fig. 4.14, arrowheads). AP492 very rarely co-localised with pericentromeric heterochromatin (data not shown). Therefore the majority of KRAB-ZFP 492 relocates from the nucleolus to the nucleoplasm upon differentiation of ES cells but the relative amount of 492 protein in the nucleus does not appear to alter. This was confirmed by western blotting (data not shown). However, the nucleolar localisation is clearly not restricted to undifferentiated cells (Fig. 4.14, differentiated SSEA-1 negative cells, with AP492 signal in nucleolus), suggesting that nucleolar distribution of KRAB-ZFP 492 may be cell-cycle/cell proliferation related. ~15% of OS25 cells, whilst undergoing selection for undifferentiation, do not stain with α-SSEA-1 (Billon *et al.*, 2002).

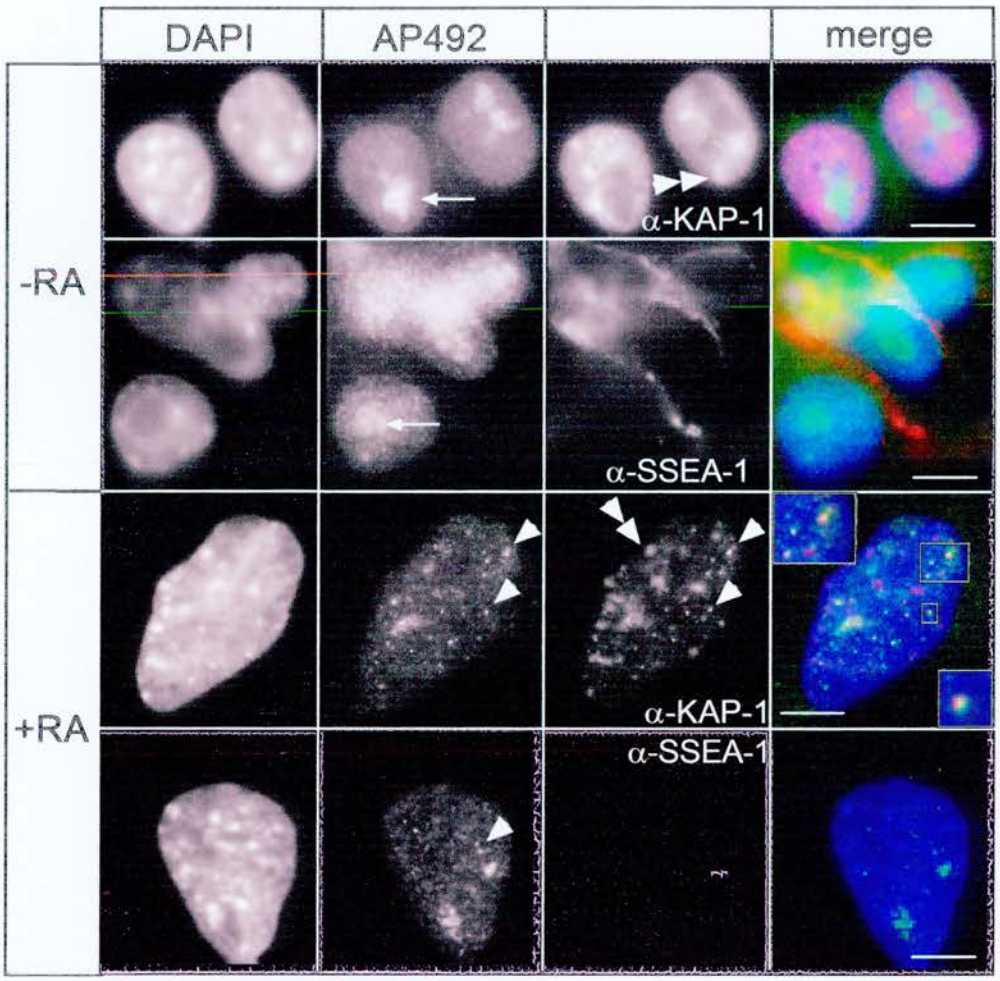


Figure 4.14 Localisation of endogeneous 492 KRAB-ZFP in OS25 cells

Immunofluorescence on undifferentiated (-RA) and differentiated (+RA) OS25 ES cells with 1:100 AP492 (green in merge) and either α-KAP-1 or α-SSEA-1 (red in merge) antibodies. Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining. Arrowheads indicate punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 mm.

Non-staining of the antibody may account for this phenomenon, or these cells may overcome the selection process and spontaneously differentiate in culture as SSEA-1 negative cells. AP492 was found to be nucleolar in both SSEA-1 positive and negative undifferentiated cells (Fig. 4.14), suggestive of a possible cell cycle regulated, nucleolar distribution of KRAB-ZFP 492.

4.9 Discussion

At the start of this thesis, the localisation of KRAB-ZFPs with respect to KAP-1 had not been extensively studied. Reports described the co-localisation of epitope-tagged KRAZ1 with KAP-1 at heterochromatin in NIH 3T3 cells (Matsuda *et al.*, 2001), and ZFP-37 (Payen *et al.*, 1998) and the gene-trapped KRAB-ZFPs (Sutherland *et al.*, 2001) were also found to be situated at heterochromatin in a subset cells. Hence I investigated the localisation of the gene-trapped KRAB-ZFPs by immunofluorescence with respect to cellular differentiation and the cell cycle. In undifferentiated gene-trap ES cells, I found that two populations of cells exist, those with and those without KRAB-ZFPs and KAP-1 concentrated at heterochromatin. Although, α-KAP-1 and α-SSEA-1 dual immunofluorescence experiments could not be performed, it is highly probable that KAP-1 is concentrated at pericentromeric heterochromatin of some undifferentiated cells. However, I found a striking re-localisation of both KAP-1 and the gene-tapped fusion proteins to pericentromeric heterochromatin upon differentiation. KAP-1 is probably required (but not sufficient) for KRAB-β-gal localisation. Concomitant with this study, Cammas *et al.* showed a significant re-localisation of KAP-1 to heterochromatin in ES and F9 cells after 48 hr of RA differentiation, but they did not examine any KRAB-ZFPs (Cammas *et al.*, 2002). Heterochromatic localisation of the gene-trapped KRAB-ZFPs, like KAP-1, was also found to be independent of the cell cycle (section 4.6; Cammas *et al*, 2002) although there appears to be a correlation with the proliferative state of cells in that cells with heterochromatic α-β-gal staining were found in BrdU negative cells more often than BrdU positive cells.

To determine if full-length proteins behave the same way as the gene-traps I made full-length 492-GFP constructs with the GFP tag in both orientations and generated various mutant 492 constructs which contained either the KRAB (+linker) or ZF domains only. With the tag in either orientation, 492-GFP localised to pericentromeric heterochromatin in ~20% of transfected cells similar to epitope tagged KRAZ1 and KRAZ2. However, the staining pattern slightly differs due to the additional appearance of smaller punctate foci that were not associated with heterochromatin. I decided to generate an antibody to the endogenous 492 gene-trap protein as it is possible that transfection of tagged proteins could lead to protein mis-localisation or aggregates of over-expressed protein. The small foci were often observed by immunofluorescence with AP/IgG492 antibodies, but interestingly, in contrast to both the gene-trap fusion protein and GFP-tagged 492, pericentromeric localisation was rarely observed. Similarly, GFP-tagged ZFP-57 was found to have large punctate foci in the nuclei of transfected NIH 3T3 cells that co-localise with HP1α in heterochromatin rich areas as revealed by Hoechst dye which identifies A/T-rich repeat sequences in heterochromatin (Alonso *et al.*, 2004). In Schwann cells, endogenous ZFP-57 antibody staining is described as small punctate foci in contrast to the larger distinct patches seen in transfected cells and this euchromatic staining also overlaps with HP1α, although these pictures are not clear in the paper (Alonso *et al.*, 2004).

In the case of the gene-traps and GFP-tagged 492, heterochromatic targeting via KAP-1 must be due to the KRAB+linker domain. Interestingly, some ESTs in the databases appear to encode KRAB-domain only containing proteins. I predict that these proteins might be distributed in cells in a similar way to the gene-traps. In support of this, a recent report describes an alternatively spliced isoform of the *Zfp208* locus that encodes a KRAB-only protein (KRAB-O). When FLAG-tagged and over-expressed in COS7 cells this construct was found to be nuclear and cytoplasmic (Oh *et al.*, 2005) similar to GFP-492K. Additionally, a punctate staining pattern in the nucleus was described with the number and size of the punctate spots being variable. Co-localisation of these large spots with heterochromatin is not obvious in this monkey cell line, but KRAB-O was shown to co-localise with FLAG-

tagged KAP-1 when both constructs were transiently transfected into COS7 cells and more cells expressed KRAB-O when co-transfected with KAP-1 suggesting over-expression of KAP-1 may stabilise the KRAB-O protein. However, endogenous antibody staining on primary cells from mouse embryonic gonads demonstrated a more granular, nuclear diffuse staining pattern that overlapped with euchromatic HP1γ proteins in a grainy, nuclear diffuse manner. Importantly, the localisation studies of KRAB-ZFP 492 in the light of this recent evidence suggests that endogenous KRAB-ZFPs may have very minor roles at pericentromeric heterochromatin, if any and the effects of epitiope tagging KRAB-ZFPs may not be the best way to study their cellular distributions. Such experiments should be interpreted with caution since distribution patterns may depend on the epitope tag, their level of expression and the cell type chosen for transfection. Whilst co-localisation studies of over-expressed KRAB-ZFPs may define interactions that are physiologically possible, they may not be mechanistically sound when compared to endogenous co-localisation studies.

Indeed, the expression level of GFP-492 varied widely among individual transfected cells, and there appeared to be a correlation between high levels of 492-GFP expression and cell death. The transfection efficiency was only ever between 10% and 20% and although it is possible that only cells in a particular stage of the cell cycle express KRAB-ZFP 492, as postulated for NRIF (Casademunt *et al.*, 1999), I propose that KRAB-ZFP 492 or its target genes, are tightly regulated by the cell and that over-expression of GFP-492 may cause excess 492 to be deposited at heterochromatin via an *in vivo* interaction with KAP-1 and not because of mis-folded protein aggregates. This over-expression may be directly or indirectly detrimental to cells, as many nuclei with heterochromatic GFP-492 staining appeared smaller and less healthy. The possible over-expression of tagged KRAB-ZFPs means that excess protein may be recruited to heterochromatin and this could be the case for KRAZ1 and KRAZ2 (Matsuda *et al.*, 2001).

Endogenous 492 KRAB-ZFP was also found in the nucleoli of both undifferentiated ES cells and NIH3T3 cells, but this staining pattern was neither observed in the gene-

trap cells nor in cells transfected with tagged 492. The nucleolar localisation of KRAB-ZFP 492 adds to the number of KRAB-ZFPs that have been found in the nucleolus (Payen *et al.*, 1998, Huang *et al.*, 1999, Yano *et al.*, 2000; and Leung *et al.*, 2003). This may reflect a nucleolar function, such as for Kid-1, which has a role in nucleolar disintegration (Huang *et al.*, 1999), or it may be one of a growing list of proteins sequestered in the nucleolus. Nucleolar localisation of ZFP-37 is restricted to certain neuronal cell types (cerebellar Purkinje cells) (Payen *et al.*, 1998) and I have shown a re-localisation of KRAB-ZFP 492 from the nucleolus during differentiation; therefore, a potential role for 492 in the nucleolus may be for storage. Proteins involved in developmentally regulated pathways, as is likely for KRAB-ZFP 492 (Chapter 6), are often compartmentalised in cells as a way of regulating their function. For example, ARL5 (a member of the ARF family of Ras-related GTPases), when in its GDP bound state, is concentrated in the nucleoli but once bound to GTP it relocates to the nucleus where it can bind to HP1α (Lin *et al.*, 2002). Recently, PML was found to regulate p53 stability by sequestering Mdm2, a negative regulator of p53, to the nucleolus (Bernardi *et al.*, 2004). KRAB-ZFP 492 may function in such way since it is probably a negative regulator and during cell differentiation could be released from the nucleolus whereupon it can bind to its target genes to repress them. This hypothesis coincides with nuclear diffuse re-localisation of KRAB-ZFP 492 and the formation of euchromatic foci in what are predominantly differentiated ES cells. How KRAB-ZFP 492 is targeted to the nucleolus is currently unknown as none of the tagged constructs showed enrichment of the protein in the nucleolus. For ZNF274, nucleolar-targeting ability is restricted to a minimal domain consisting of the third and fourth ZF (Yano *et al.*, 2000). The presence of 492 in the nucleolus may depend on its association with RNA and/or DNA or other nucleolar and/or nuclear proteins, and such interactions may be affected by protein modifications since nucleolar localisation of PML is dependent on ATR activation and subsequent phosphorylation of PML by ATR (Bernardi *et al.*, 2004).

Recent evidence suggests that although the same domains are contained within KRAB-ZFPs, they are not necessarily responsible for the same cellular targeting and

this has vital functional implications. Construction of mutant/truncated versions of some KRAB-ZFPs and their subsequent epitope tagging has elucidated the domains, at least in part, involved in cellular targeting (Grondin *et al.*, 1996; Yano *et al.*, 2000; Jheon *et al.*, 2001; Mark *et al.*, 2001; Sun *et al.*, 2003; Hennemann *et al.*, 2003; Gou *et al.*, 2004; Alonso *et al.*, 2004). For 492,the gene-trapped protein suggests that the KRAB (+linker) domain is exclusively responsible for nuclear targeting, but GFP-492K shows cytoplasmic staining also. Similarly, the KRAB domain of ZNF300 is diffuse throughout the whole of the cell whilst the ZF domain is targeted to the nucleus (Gou *et al.*, 2004). In both cases, cytoplasmic staining may be the result of over-expression. The ZF domain of 492 is not responsible for nuclear targeting and none of the 492 constructs were responsible for enrichment of the protein in nucleoli. Other protein-binding partners of KRAB-ZFP 492 may therefore be responsible for its localisation, as is the case for NRIF (Gentry *et al.*, 2004). C-terminal mutants of NRIF (lacking ZF domain), which were able to associate with TRAF6 (a member of the TNF receptor subfamily), were never observed within the nucleus, but accumulated in perinuclear regions. It is speculated that the KRAB domain of NRIF prevents its entry into the nucleus and that this block is only relieved upon binding to TRAF6, although how TRAF6 allows this nuclear targeting is unknown (Gentry *et al.*, 2004). The N-terminal mutants that were unable to bind TRAF6 were found exclusively in the nucleus. Therefore the binding of these two proteins is required for changes in sub-cellular distribution of NRIF suggesting that KRAB-ZFPs can shuttle between the nucleus and the cytoplasm depending on protein-protein interactions (Gentry *et al.*, 2004). Endogenous 492 localisation in the nucleus and nucleolus provides further evidence for KRAB-ZFP shuttling.

Matsuda *et al.* (2001) suggest that pericentromeric localisation of KRAZ1 and KRAZ2 is central to their ability to repress transcription. Based on the results of this chapter, my supposition is that KRAB-ZFPs need not be recruited to pericentromeric heterochromatin for KAP-1 mediated repression. Furthermore, it is perhaps at the smaller foci, which form on differentiation, that repression of a specific or set of target genes may occur since KAP-1 was also localised at some of these foci. Since KRAB-ZFP 492 was rarely enriched at heterochromatin, it is possible that the

KRAB-KAP-1 interaction at heterochromatin is a transient one. On this note, KRAB-ZFP 492 was never completely excluded from heterochromatin. The question remains therefore, as to why KAP-1 is enriched at heterochromatin upon differentiation if not to mediate gene repression via the KRAB domain? The report of a novel KAP-1-associated histone H3, lysine 9 (H3-K9) specific methyltransferase, SETDB1, which was shown to contribute to KAP-1-HP1-mediated silencing of euchromatic genes (Schultz *et al.*, 2002) influenced me to pursue the small punctate foci that I observed for KRAB-ZFP 492, with a view to further understanding the role of KRAB-ZFPs in KAP-1 mediated repression.

Chapter 5

Investigation into the role of KRAB-ZFPs in KAP-1 mediated repression

5.1 Introduction

In the last two years, only a few papers have built on the initial KRAB-KAP-1 mediated repression hypothesis (Schultz *et al.*, 2002; Cammas *et al.*, 2002; Ayyanathan *et al.*, 2003; Cammas *et al.*, 2004) (Fig. 1.6 and 5.1AI). The first of the KRAB-KAP-1 mediated repression models suggests that regulation of cell-type specific genes or sub-sets of genes could be a direct result of KRAB-ZFPs binding to target promoters, with subsequent translocation of the gene to centromeric heterochromatin via the KRAB-KAP-1-HP1 interaction (Fig. 1.6 and 5.1AI). Evidence for this model is suggested from the finding that the PxVxL motif of KAP-1 is essential for both HP1 interaction, centromeric localisation, and thus gene silencing (Matsuda *et al.*, 2001; Cammas *et al.*, 2002). Whilst this interaction and dynamic re-location to heterochromatin of KAP-1 is not required for initial RA differentiation of F9 EC cells into primitive endoderm cells (PrE), it is required for differentiation into visceral endoderm (VE) and parietal endoderm-like (PE) on addition of cAMP (cyclic AMP) (Cammas *et al.*, 2004). The KAP-1-HP1 interaction is required during a short window of time within early differentiating PrE cells (between 1-2 days of RA treatment) for their terminal differentiation into PE and VE and the necessity of the KAP-1-HP1 interaction during this time frame coincides precisely with re-localisation of KAP-1 from eu- to heterochromatin. The consequent down-regulation of 3 endoderm-specific genes (*GATA6*, *HNF4* and *Dab2*) during PrE differentiation of cells with a disrupted KAP-1-HP1 interaction lends support to the idea that the KAP-1-HP1 co-repressor complex may trigger the translocation of their direct target genes from eu- to heterochromatin for their silencing in order to allow the normal expression of *GATA6*, *HNF4* and *Dab2*. That KRAB-ZFPs may relocate their target genes to heterochromatin for silencing is based on the model of Ikaros-mediated silencing of the *Tdt* genes during T-cell activation (Brown *et al.*, 1997; Brown *et al.*, 1999) and is suggested by the observation that two KRAB-ZFPs (KRAZ1 and KRAZ2) co-localise with KAP-1 and HP1 proteins within pericentromeric heterochromatin (Matsuda *et al.*, 2001). In support of this model, a euchromatic luciferase transgene has recently been shown to

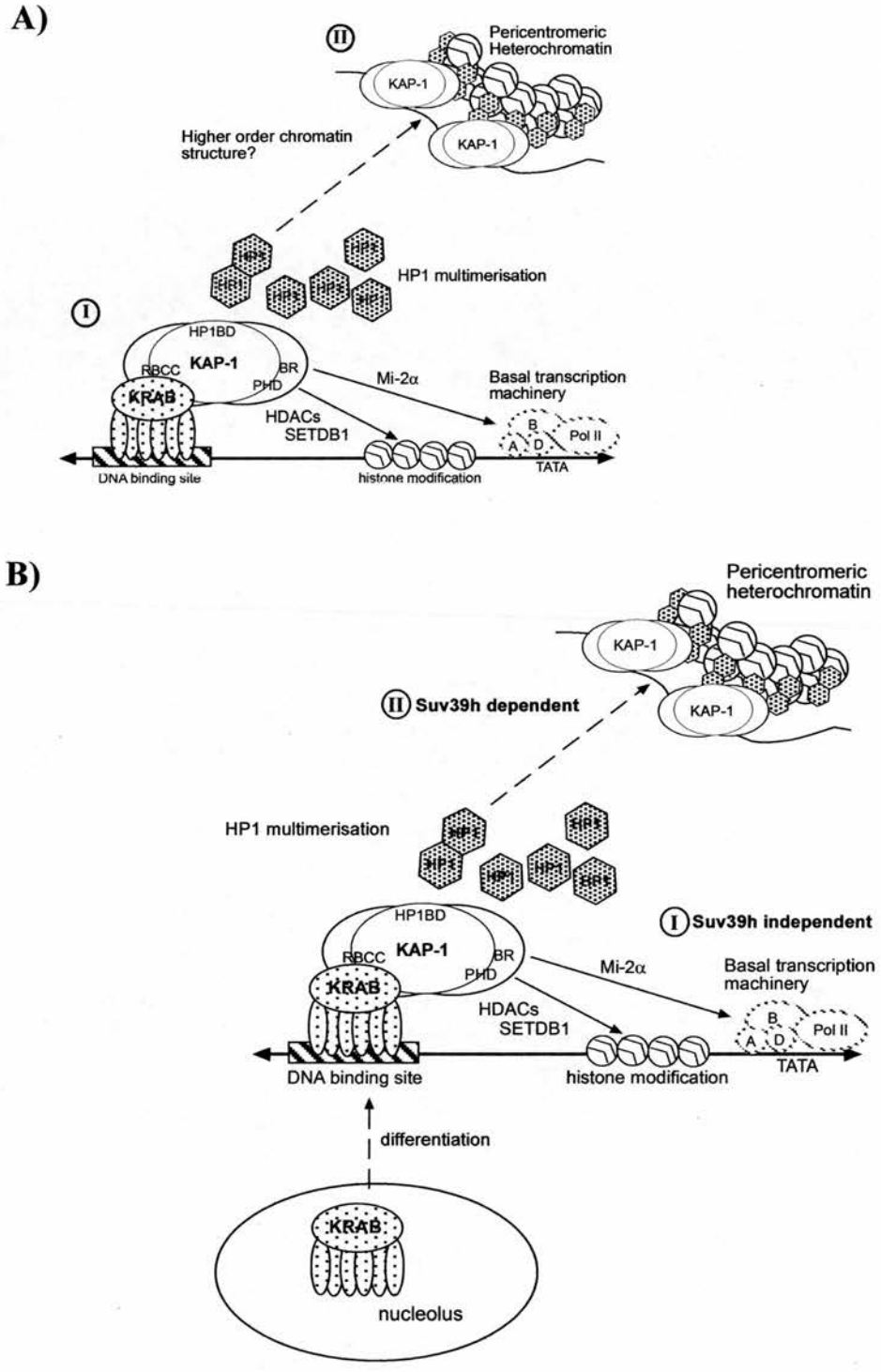


Figure 5.1 The KRAB-KAP-1 repression model

A) Proposed model for KRAB-KAP-1 mediated repression. Refer to Figure 1.6 for legend. **B)** KRAB-ZFPs are stored in the nucleolus in undifferentiated cells. Upon differentiation, KRAB-ZFPs are released from the nucleolus were they can bind to their specific target genes. (I) Localised repression of KRAB-ZFP genes, mediated by KAP-1, occurs in KRAB and KAP-1 associated (KAKA) foci and is Suv39h independent. (II) KAP-1 recruitment to heterochromatin is Suv39 dependent.

re-locate to heterochromatin upon repression by a hormone-regulatable DNA-binding KRAB-domain fusion protein (Ayyanathan *et al.*, 2003).

The second proposed model is that individual KRAB-ZFPs target specific genes for silencing by recruitment of KAP-1, whose oligomerisation could form the basis of a molecular platform to co-ordinate the subsequent recruitment of the NuRD HDAC complex, SETDB1 and HP1 (Fig. 1.6 and 5.1AII). This would facilitate the nucleation of facultative heterochromatin at the locus (Schultz *et al.*, 2002; Ayyanathan *et al.*, 2003). The formation of a local heterochromatic environment is based on evidence, initially uncovered in a yeast two-hybrid assay, that showed the interaction of a novel SET-domain protein, SETDB1 with the PHD and bromodomain of KAP-1 (Schultz *et al.*, 2002). A different region of this interaction surface was previously found to be responsible for interaction with Mi-2α (a novel subunit of the NuRD HDAC complex, see section 1.3.3). The SETDB1 H3-K9 histone methyltransferase (HMT) localises exclusively in euchromatic regions of NIH 3T3 cells and overlaps with euchromatic HP1α staining (Schultz *et al.*, 2002). The study of KAP-1, SETDB1 and HP1 recruitment to euchromatic genes targeted by KRAB-ZFPs then ensued with a series of experiments on NIH 3T3 cells comprising a hormone inducible two-plasmid system (Schultz *et al.*, 2002; Ayyanathan *et al.*, 2003). This system comprises a reporter plasmid that has been stably integrated into a highly transcribed region of euchromatin, and an expression plasmid encoding a hormone-regulatable DNA-binding KRAB-domain fusion protein. The transient and reversible targeting of endogenous KAP-1 and its associated factors to the *tk* promoter and its effects on luciferase transgene expression in the presence/absence of oestrogen (KRAB-fusion protein binding does not occur in the absence of oestrogen) was analysed. ChIP (chromatin immunoprecipitation) analysis showed enrichment of SETDB1, KAP-1, HP1α and H3-methylated K9 at the *tk* promoter upon oestrogen induction but HP1γ remained constitutively bound. These chromatin changes were highly localised to the luciferase transgene and did not spread to the linked promoters of the selection cassettes (zeocin expression from the linked zeocin cassette ~2.8 kb away was unaffected). Localised chromatin compaction in the promoter region was

demonstrated by lack of nuclease accessibility. Similar results were achieved with ChIP on an endogenous target gene, *Col11α2*, using antibodies to its cognate KRAB-ZFP, NT-2, although importantly, whether the target gene re-localised to pericentromeric heterochromatin was not analysed in this study (Ayyanathan *et al.*, 2003).

Whilst the above models are not mutually exclusive, both can be corroborated by the sub-nuclear distribution of KAP-1, neither has yet been proven by a thorough study of localisation of an endogenous KRAB-ZFP or a target gene, due to the scarcity of both KRAB-ZFP antibodies or known targets. In this chapter I have attempted to build on the two proposed models of KRAB-KAP-1 mediated repression with the reagents available. To this end, I have characterised the euchromatic foci of KRAB-ZFP 492 (Chapter 4) and have tried to address whether these foci are specific to 492 or whether they are a general frequent of KRAB-ZFPs by investigating the sub-nuclear distribution of another KRAB-ZFP, NT-2. Since SETDB1 is reported to be euchromatic, I surmise that it is not involved in localisation of KAP-1 to pericentromeric heterochromatin and have therefore examined whether or not the Suv39h1 and Suv39h2 H3-K9 HMTs, which direct DNA methylation to major and minor satellite repeats at pericentromeric heterochromatin (Lehnertz *et al*, 2003), play a role here instead by investigating what happens to KAP-1 and the 492 foci in *Suv39h* dn cells. Since DNA methylation might play a role in the mitotically heritable gene silencing of the luciferase transgene (Ayyanathan *et al.*, 2003), I have also examined the sub-nuclear distribution of KAP-1 and the 492 foci in Dnmt3ab$^{-/-}$ cells, which lack any DNA methylation (Jackson *et al.*, 2004).

5.2 Euchromatic foci of KRAB-ZFP 492 co-localise with HP1α and HP1β

I think it most likely that the discrete 492 euchromatic foci (section 4.8.5) correspond to small territories, or islands, of facultative heterochromatin as predicted by the latter of the two models described above. These foci could represent the capacity of KAP-1 to co-ordinate the biochemical activities required to induce and maintain the

assembly of a local, but higher-order, chromatin structure at a genomic region targeted by a given KRAB-ZFP. I therefore decided to test for the presence of the HP1 proteins within these discrete foci.

In mammalian cells, the HP1 proteins have specific nuclear staining patterns depending on the cell type and the level of expression. HP1α is mostly concentrated at sites of constitutive heterochromatin in both human HT1080 cells (Gilbert *et al.*, 2003), and murine cells, for example, NIH 3T3, C127 and EC cells (Horsley *et al.*, 1996; Neilsen *et al.* 1999; Minc *et al.* 2000; Cammas *et al.*, 2002). HP1β and γ are also detectable at sites of pericentromeric heterochromatin in mouse cells, but there is also a large pool of these proteins distributed more diffusely throughout the nucleoplasm. HP1γ is predominantly found at euchromatic sites (Horsley et al., 1996; Neilsen *et al.*, 1999). In contrast, HP1β and γ cannot be seen associated with centromeric domains in human cells but there are clear non-centromeric foci, which most likely correspond to other heterochromatin domains (Gilbert *et al.*, 2003). In mouse elongating spermatids HP1α is expressed at very low levels and goes undetected by immunofluorescence, whereas HP1β and γ both exhibit a predominant localisation to the chromocenter and other additional sites dispersed within the nucleoplasm of this cell type (Khetchoumian *et al.*, 2004). Notably, KAP-1 is not expressed in elongating spermatids (Weber *et al.*, 2002) although TIF1δ is, and co-localises with HP1γ at these additional nucleoplasmic sites that are postulated to be heterochromatin or heterochromatin-like repressive structures (Khetchoumian *et al.*, 2004). This shows that the HP1 proteins have distinct euchromatic roles in addition to their function at pericentromeric heterochromatin.

Undifferentiated and differentiated OS25 cells and NIH 3T3 were fixed with 3% pFa onto slides for immunofluorescence (section 2.8.1). Cells were incubated with AP492 and either α-HP1α, α-HP1β or α-HP1γ antibodies. Dual HP1 and KAP-1 immunofluorescence could not be performed because all antibodies were raised in mouse. HP1 sub-nuclear localisation in OS25 cells was comparable with that reported in other mouse cell types. As shown previously in undifferentiated OS25 cells, 492 was enriched in nucleoli and also showed diffuse nucleoplasmic staining.

However, in a small proportion of NIH 3T3 cells, and more particularly in differentiated OS25 cells (~5-10%), 492 foci were observed. These were found to be juxta-positioned to both non-pericentromeric foci of both HP1α and HP1β (Fig. 5.2, arrowheads). Confocal imaging confirmed these staining patterns with HP1α (Fig. 5.3, arrowheads). KRAB-ZFP 492 foci were not always adjacent to HP1 (Fig 5.3) and/or KAP-1 foci (Fig. 4.14) verifying that merged signal is not a result of bleed-through fluorescence. Nuclei containing more prominent 492/HP1 foci were often larger and flatter than other differentiated cells on the same slide (Figures 4.12, 4.14 and 5.2). Co-localisation of foci with HP1γ was not observed (Fig. 5.2). This may be due to the stronger euchromatic staining observed with this antibody so that defined and discrete foci cannot often be visualised. However, enrichment of HP1γ also does not occur in ChIP experiments with the integrated luciferase transgene (Ayyanathan *et al.*, 2003).

The foci of staining with AP492 might represent a novel nuclear domain or co-localisation to a known nuclear compartment. Therefore I performed dual immunofluorescence with antibodies to PML (promyelotic leukaemia) nuclear bodies and CREST anti-sera, to see if these foci marked either of these nuclear sub-compartments. Centromeres are composed of both major and minor satellite DNA, the latter of which form distinct sub-nuclear compartments not necessarily represented by the DAPI bright spots of associated major satellites, and are associated with different sets of non-histone proteins (Guenatri *et al.*, 2004). The well-characterised nuclear dot-like structures previously described for PML, SC35 and BRCA1 had already been tested for co-localisation with KAP-1, but the emphasis of co-localisation was with respect to the large foci of pericentromeric KAP-1 staining (Ryan *et al.*, 1999). PML has been implicated as a cofactor of TIF1α-dependent enhancement of RARα-mediated transcriptional activation (Zhong *et al.*, 1999) and a splice variant of a major constituent of PML bodies, SP100C, has homology to the TIF1 proteins, with all homologues being covalently modified by SUMO (Seeler *et al.*, 2001). SP100 is also known to bind to HP1 proteins (Seeler *et al.*, 1998). It is therefore conceivable that PML bodies may also contain KRAB-ZFPs via an indirect or transient association with TIF1α and/or KAP-1. However,

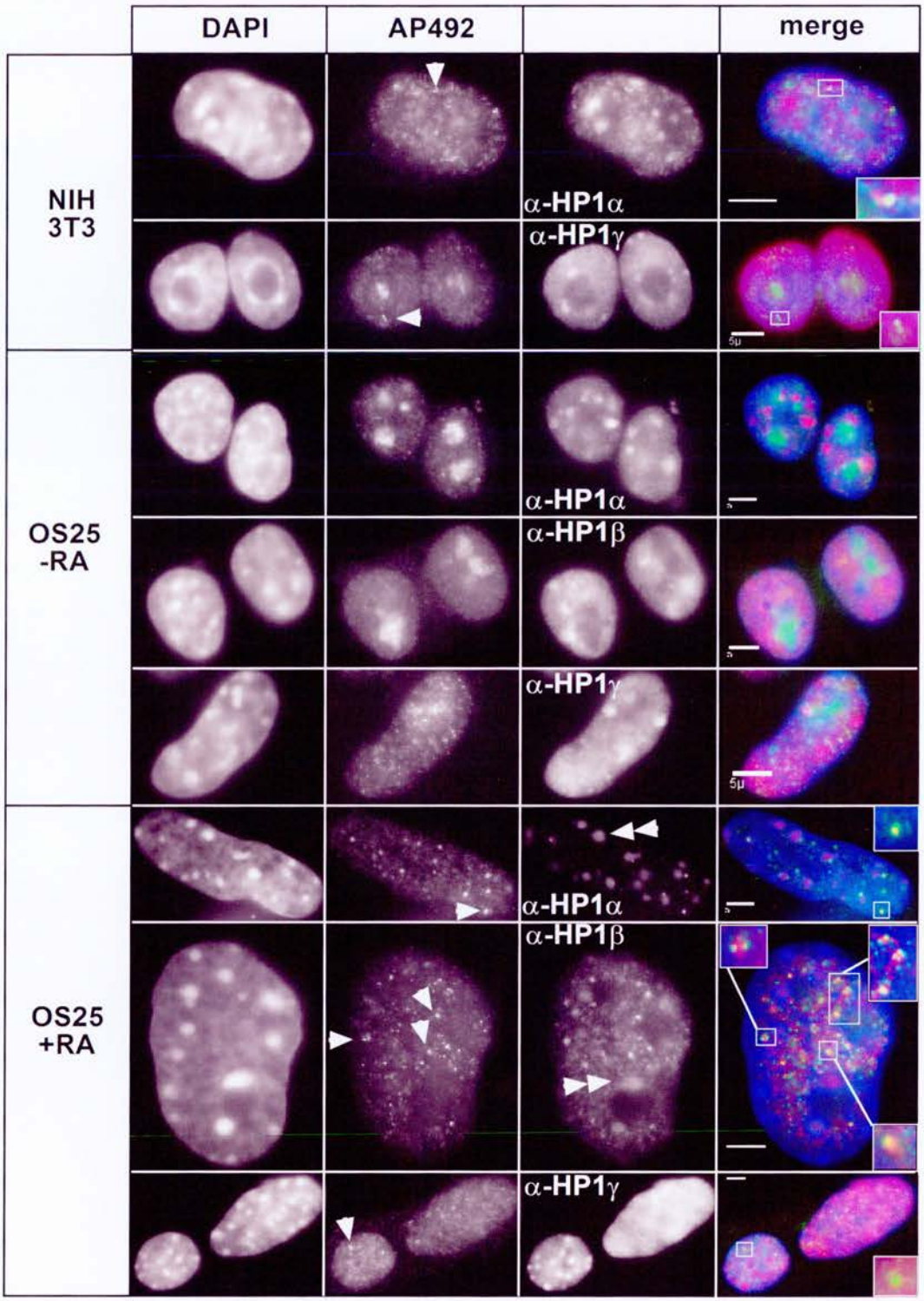


Figure 5.2 Localisation of endogenous 492 KRAB-ZFP and the HP1 isoforms
Immunofluorecence of NIH 3T3 and undifferentiated and differentiated OS25 ES cells with 1:100 AP492 (green in merge) and α-HP1 (red in merge) antibodies. Nuclei are counterstained with DAPI (blue in merge). Arrowheads indicate 492 punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 μm.

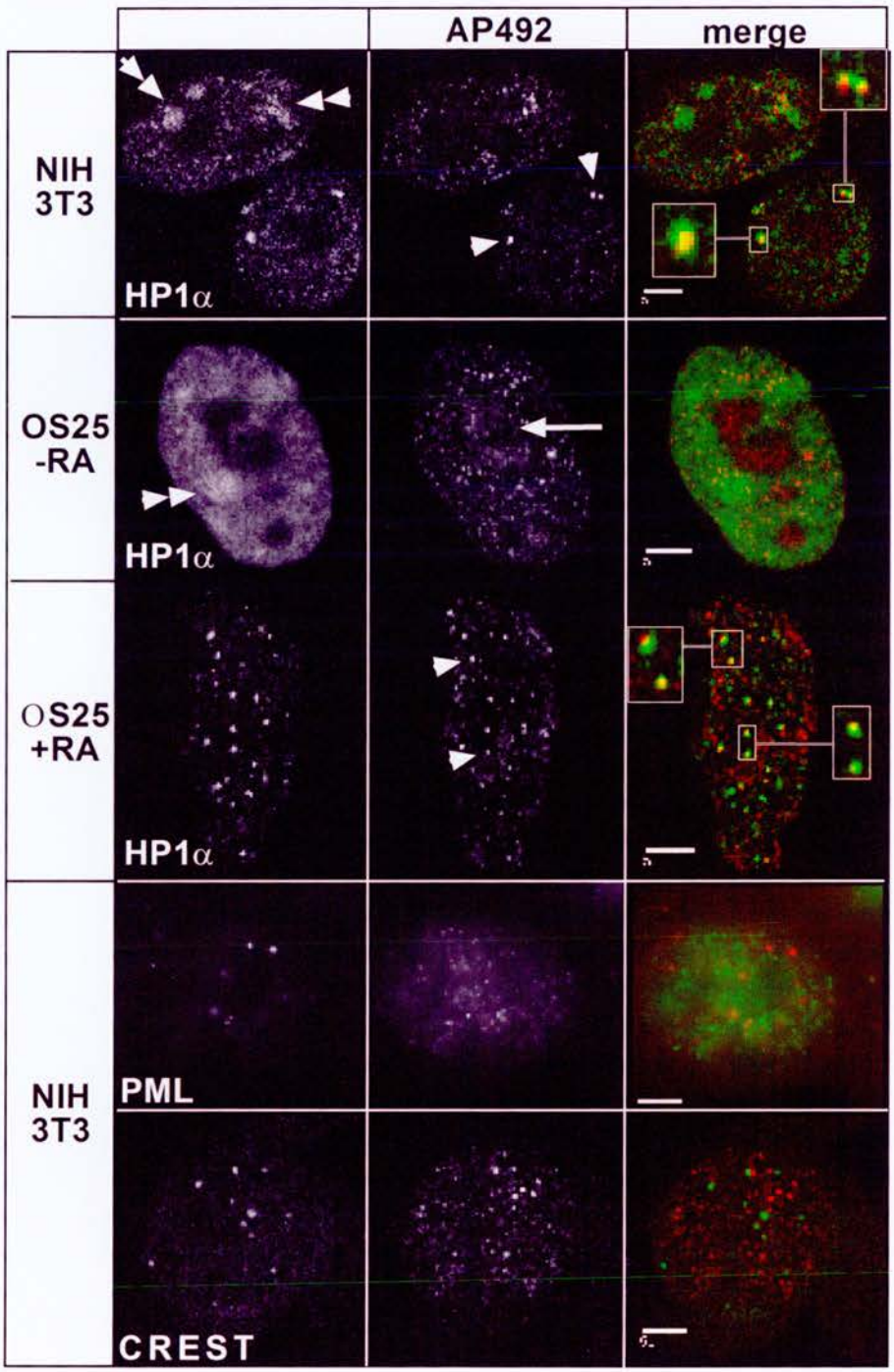


Figure 5.3 Confocal microscopy of KRAB-ZFP 492 foci
Immunofluorecence of NIH 3T3, undifferentiated (-RA) and differentiated (+RA) OS25 ES cells with AP492 (green in merge) and either α-HP1α, α-PML or human CREST α-sera. A confocal image is not shown for AP492 and α-PML co-staining. Arrows indicate cells with nucleolar staining. Arrowheads indicate 492 punctate foci and double arrowheads indicate heterochromatic staining. Bars= 5 μm.

immunofluorescence with AP492 and either α-PML antibodies or human CREST anti-sera with subsequent fluorescent/confocal imaging demonstrated that the 492 foci did not significantly overlap with either of these nuclear sub-compartments (Fig. 5.3).

The KRAB-ZFP 492 foci could represent regions occupied by target genes, or they could be a consequence of other protein-protein interactions mediated by the ZF or linker region. Various KRAB-ZFPs are known to interact with other proteins via their ZF, KRAB or linker regions (section 7.6). But whilst homo-oligomerisation of ZBRK1 and NRIF has been reported (Tan *et al.*, 2004a; Gentry *et al.*, 2004), there is no evidence to suggest that different KRAB-ZFPs can directly interact. Hence the KRAB-ZFP 492 foci may represent a novel sub-nuclear compartment in which other KRAB-ZFPs may also be present and this nuclear organisation may help them to achieve their common function of gene silencing. The co-localisation of transcribed genes to functional sub-nuclear compartments called 'transcription factories' for gene transcription has recently been described (Osborne *et al.*, 2004). In pursuit of whether the 492 foci represent the assembly of a novel repression complex at a euchromatic gene or whether they represent a 'silencing factory' to where different KRAB-ZFP target genes migrate upon regulation by a number of individual KRAB-ZFPs, I decided to investigate the sub-nuclear localisation pattern of the only other murine KRAB-ZFP, NT-2, for which an antibody was available at that time (Tanaka *et al.*, 2002), and for which the nuclear localisation had not yet been reported.

5.3 Analysis of the sub-nuclear localisation of another KRAB-ZFP, NT-2

Until October 2003, NT-2 was the **only murine KRAB-ZFP** for which a physiological target gene was known. NT-2 is a SCAN domain-containing KRAB-ZFP and binds in a 24 bp region of the *Col11α2* promoter. The role of *NT-2* as a negative regulator of the cartilage specific *Col11a2* (α2[XI] collagen) gene in a differentiating cell system has been documented; and during differentiation of the murine EC-derived chondrogenic cell line, ATDC5, *NT-2* mRNA levels are inversely

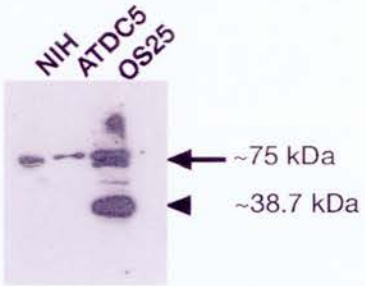
correlated with those of *Col11α* (Tanaka *et al.*, 2002). *NT-2* mRNA is abundant in undifferentiated ATDC5 cells, but decreases when the cells differentiate into proliferative chondrocytes (d6-d15) and begin to express *Col11α2* (Fig. 5.4A). Further culturing into hypertrophic chondrocytes (d15-d21), concomitant with the formation of cartilage nodules, up-regulates *NT-2* mRNA expression and expression of *Col11α2* is repressed with the onset of *Col10α1* expression. Thanks to Dr Yoshihiko Yamada, who kindly provided NT-2 antibody, I was able to examine the localisation of NT-2 by immunofluorescence during the differentiation of ATDC5 cells.

Firstly, I analysed the NT-2 antibody by western blotting. As previously described, NT-2 antibody detects a prominent band of ~75 kDa in NIH 3T3 and ATDC5 cells (Fig. 5.4B), although lower MW bands could also be identified after longer exposures (data not shown) (Tanaka *et al.*, 2002; Ayyanathan *et al.*, 2003). In OS25 cell extracts additional bands were detected including a very prominent ~38 kDa species (Fig. 5.4B). These bands could be non-specific, or alternative isoforms of the NT-2 KRAB-ZFP. Despite the concern of specificity, I tested the α-NT-2 antibody by immunofluorescence in NIH 3T3 and in undifferentiated and differentiated OS25 cells. ~30-50% of NIH 3T3 cells showed nucleolar localisation of NT-2 which co-localised with fibrillarin (data not shown). The remaining cells showed a nuclear diffuse localisation with discrete foci in ~10% of cells (Fig. 5.4C, arrowheads), but there was no co-localisation with KAP-1 (data not shown) or HP1α (Fig. 5.4C). Dual immunofluorescence in undifferentiated OS25 cells with α-NT-2 and α-fibrillarin showed NT-2 to be in one or more nucleoli or a patch adjacent to one, but never the largest nucleoli (Fig. 5.4C, arrows) and often no co-localisation was seen at all (data not shown). However, re-distribution of NT-2 from 'minor nucleoli/patches' in undifferentiated OS25 cells (Fig. 5.4C, arrows) to the nucleoplasm was observed upon differentiation. This is reminiscent of 492 re-distribution. Foci were observed in differentiated OS25 cells, but again no obvious co-localisation with HP1α or with KAP-1 was observed (Fig. 5.4C, arrowheads). No co-localisation was observed with the other HP1 isoforms (data not shown). I therefore decided to study the localisation of KAP-1 and NT-2 as it regulates the

A)

	NT-2	*Col11α2*	*Col10α1*
undiff	↑	↓	X
d6-d15	↓	↑	X
d15-d21	↑	X	↑

B)



C)

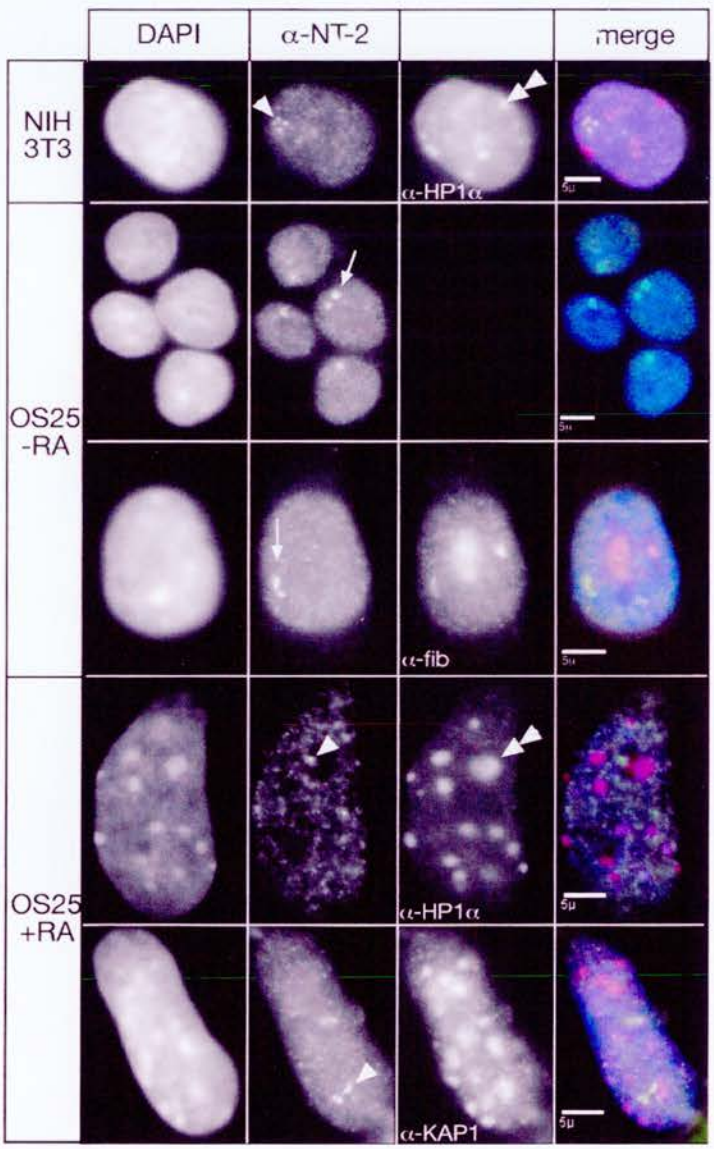


Figure 5.4 Analysis of NT-2 antibodies by western blotting and immunofluorescence

(**A**) Schematic to show *NT-2* and *Col11α2* mRNA levels in differentiating ATDC5 cells. *X*, undetected. (**B**) Western blots of NIH 3T3, ATDC5 and wt OS25 whole cell extracts with α-NT-2 antibody (1:1250). An arrow indicates the prominant ~75 kDa band after a 45 sec exposure. (**C**) Immunofluorescence of NIH 3T3 and OS25 ES cells before (-RA) and after differentiation (+RA) with 1:100 α-NT-2 (green in merge) and either α-HP1α or α-fibrillarin (α-fib) (red in merge) antibodies. Nuclei are counterstained with DAPI (blue in merge). Arrows indicate α-NT-2 localisation in an unknown domain in undifferentiated OS25 cells. Arrowheads indicate punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 μm.

expression of *Col11α2* during ATDC5 cell differentiation. I differentiated ATDC5 cells for 26 days as previously described (Shukunami *et al.*, 1996) (section 2.7.3) and assayed the differentiation status of the cells at days 6, 10, 14, 21 and 26 of differentiation by staining in 0.1% alcian blue, which stains the extra-cellular cartilage matrix. Small amounts of alcian blue positive cartilage formation were first detectable at day 6 of differentiation after which they increased in size due to maturation of cells and accumulation of collagen matrix (Shukunami *et al.*, 1996) (Fig. 5.5A). Immunofluorescence on pFa fixed cells with α-NT-2 showed that in undifferentiated ATDC5 cells NT-2 is diffusely localised within the nucleoplasm, but ~50% cells showed nucleolar staining that co-localised with α-fibrillarin (Fig.5.5B). KAP-1 has a diffuse nucleoplasmic distribution at this time point. From day 6 onwards, a proportion of ATDC5 nuclei were smaller and more rounded compared to control nuclei, which reflects the time when ATDC5 cells are known to undergo a transient condensation of cells before the formation of nodules (Shukunami *et al.*, 1996). Between day 6-8, NT-2 has a granular nuclear diffuse staining pattern that appears gradually less intense in some cells, consistent with NT-2 down-regulation in nodule forming cells where *Col11α2* expression is high (data not shown) but nucleolar staining is still present in some cells at day 7 of differentiation (data not shown) and no obvious foci could be seen. During this time frame, KAP-1 remained dispersed throughout the nucleoplasm in both nodule and non-nodule forming cells, with no localisation to pericentromeric heterochromatin (Fig. 5.5B). However, HP1α was at pericentromeric heterochromatin in most cells. Between days 9-15, NT-2 staining became more intense and punctate. Some cells showed adjacent/overlapping NT-2/KAP-1 signal, reminiscent of the KRAB-ZFP 492 foci (Fig. 5.5B, arrowheads). It is not known whether these cells represent nodule or non-nodule forming cells. The appearance of these foci may correlate with transition from *Col11α2* expression to *Col10α1* expression during nodule formation but they may also represent cells that have already undergone the transition when *Col11α2* is being repressed. At day 15, HP1α remained predominantly localised to pericentromeric heterochromatin (Fig. 5.5B). NT-2 was still nucleolar in a proportion of cells (data not shown) but this was difficult to quantify because most of

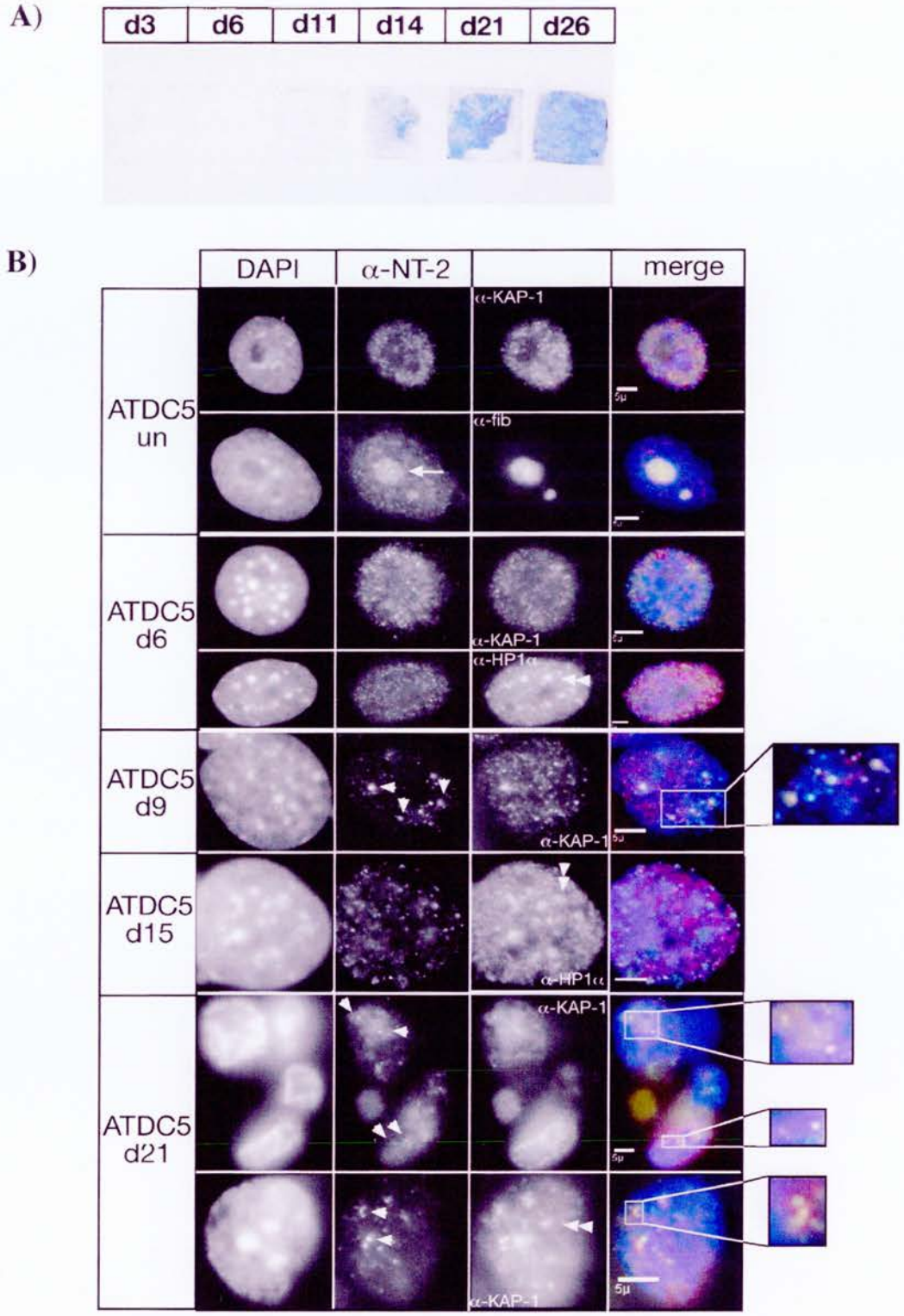


Figure 5.5 Analysis of sub-nuclear distribution of NT-2 during ATDC5 differentiation

(**A**) The differentiation status of ATDC5 cells was monitored by 0.1% alcian blue staining throughout the 26 day (d) differentiation procedure. (**B**) Immunofluorescence of undifferentiated and differentiatedATDC5 cells with 1:100 α-NT-2 (green in merge) and either α-KAP-1, α-fibrillarin (α-fib) or α-HP1α antibodies (red in merge). Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining. Arrowheads indicate punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 μm.

the cells were nodule forming and growing in polylayers due to the loss of contact inhibition (Shukunami *et al.*, 1996). At day 21-26, cells were smeared onto slides before fixation and allowed to air dry so that antibodies could have better access to the cells with as little disruption to the immediate cellular environment as possible. The DNA in these cells is considerably condensed at this time point and a layer of heterochromatin, as visualised with DAPI staining, is observed surrounding the nuclear periphery of many cells. In ~30-40% cells KAP-1 was enriched at this perinuclear heterochromatin (Fig. 5.4B). In many cells KAP-1 and NT-2 are in adjacent/overlapping euchromatic foci (Fig. 5.5B), which often appear close to or in nucleoli, but the nucleolus is not an obvious structure at day 21 and dual immunofluorescence experiments were not performed with α-fibrillarin at day 21.

NT-2, similar to KRAB-ZFP 492, appears to be re-distributed from the nucleolus to the nucleoplasm upon differentiation. It seems plausible that these two KRAB-ZFPs, could either be stored in the nucleolus before the onset of differentiation, or at particular stages of the cell cycle, or that they may even carry out a function there such as RNA binding. In some cells KAP-1 appears to be enriched at heterochromatin upon ATDC5 cell differentiation into hypertrophic chondrocytes but I did not test HP1α at this stage of differentiation. I predict that it is likely to be co-localised with KAP-1 in these cells unless the HP1 interaction is not responsible for KAP-1 localisation in this cell type.

5.4 Investigation into KAP-1 recruitment to pericentromeric heterochromatin in ES cells

To further understand the conditions of KAP-1 translocation to pericentromeric heterochromatin and the formation of 492 foci I decided to investigate this in both *Suv39h* dn cells and *Dnmt3a/b*$^{-/-}$ (*Dnmt3a* and *3b* double knockout) cells. The Suv39h1 and h2 (Suv39h) HMTs are enriched at pericentromeric heterochromatin (Aagaard *et al.*, 1999) and selectively tri-methylate histone H3 at K9 to create a high-affinity binding site for HP1 (Lachner *et al.*, 2001; Bannister *et al.*, 2001). They therefore play a central role in gene silencing at heterochromatin (Schotta *et al.*,

2002). In *Suv39h* dn cells, lack of tri-methylated H3-K9 causes HP1α and HP1β to be de-localised from pericentromeric heterochromatin (Peters *et al.*, 2001; Lehnertz *et al*, 2003). Since HP1 interaction is known to be vital for KAP-1 translocation (Cammas *et al.*, 2004), I hypothesised that KAP-1 translocation would be lost in *Suv39h* dn cells. Dnmt3a and 3b also localise to pericentromeric heterochromatin in ES cells (Bachman *et al.*, 2001) via their PWWP domain (Chen *et al.*, 2004) and Dnmt3a interacts, both *in vitro* and *in vivo*, with Suv39h1 via its PHD-like motif (Fuks *et al.*, 2003). HP1β, a physical and functional partner of Suv39h1, also binds directly to Dnmt3a (Fuks *et al.*, 2003). Moreover, in *Suv39h* dn cells, Dnmt3b fails to localise to pericentromeric heterochromatin. However, H3-K9 trimethylation is not impaired in *Dnmt3a/b*$^{-/-}$ cells (Lehnertz *et al.*, 2003). That both of these enzymatic activities might be required for KRAB-KAP-1 mediated repression was hinted at in recent experiments by Ayyanathan *et al.* in 2003 when increased DNA methylation of the *tk* promoter region (also enriched in KAP-1, HP1 and SETDB1 proteins) was found in stably silenced luciferase clones. Re-activation of the silent locus by the combined action of 5-azacytidine (5-AZ) and TSA, to relieve both histone methylation and histone deacetylation respectively, suggests that H3-K9 methylation and HP1 recruitment on a euchromatic gene may ultimately lead to DNA methylation.

I first confirmed that pericentromeric heterochromatin is enriched for H3-K9 tri-methylation and HP1 localisation in wt ES cells, from which the *Suv39h* dn cell line was generated (Table 2.6). I then performed immunofluorescence on undifferentiated and differentiated *Suv39h* dn cells with AP492 and α-KAP-1 antibodies (Fig. 5.6). In undifferentiated wt and *Suv39h* dn cells, 492 was nucleolar in a similar proportion of cells to that reported for OS25 cells (Fig. 5.6, arrow). As expected, HP1α and KAP-1 were not concentrated at heterochromatin in differentiated *Suv39h* dn cells supporting the idea that tri-methylation of H3-K9 at major satellite DNA via HP1 is required for KAP-1 targeting to heterochromatin (Cammas *et al.*, 2004). However, in differentiated *Suv39h* dn cells discrete 492 foci were still observed, similar to wt cells, (Fig. 5.6, arrowheads) and these co-localised with both KAP-1 and HP1 proteins suggesting that Suv39h1 and h2 are not required

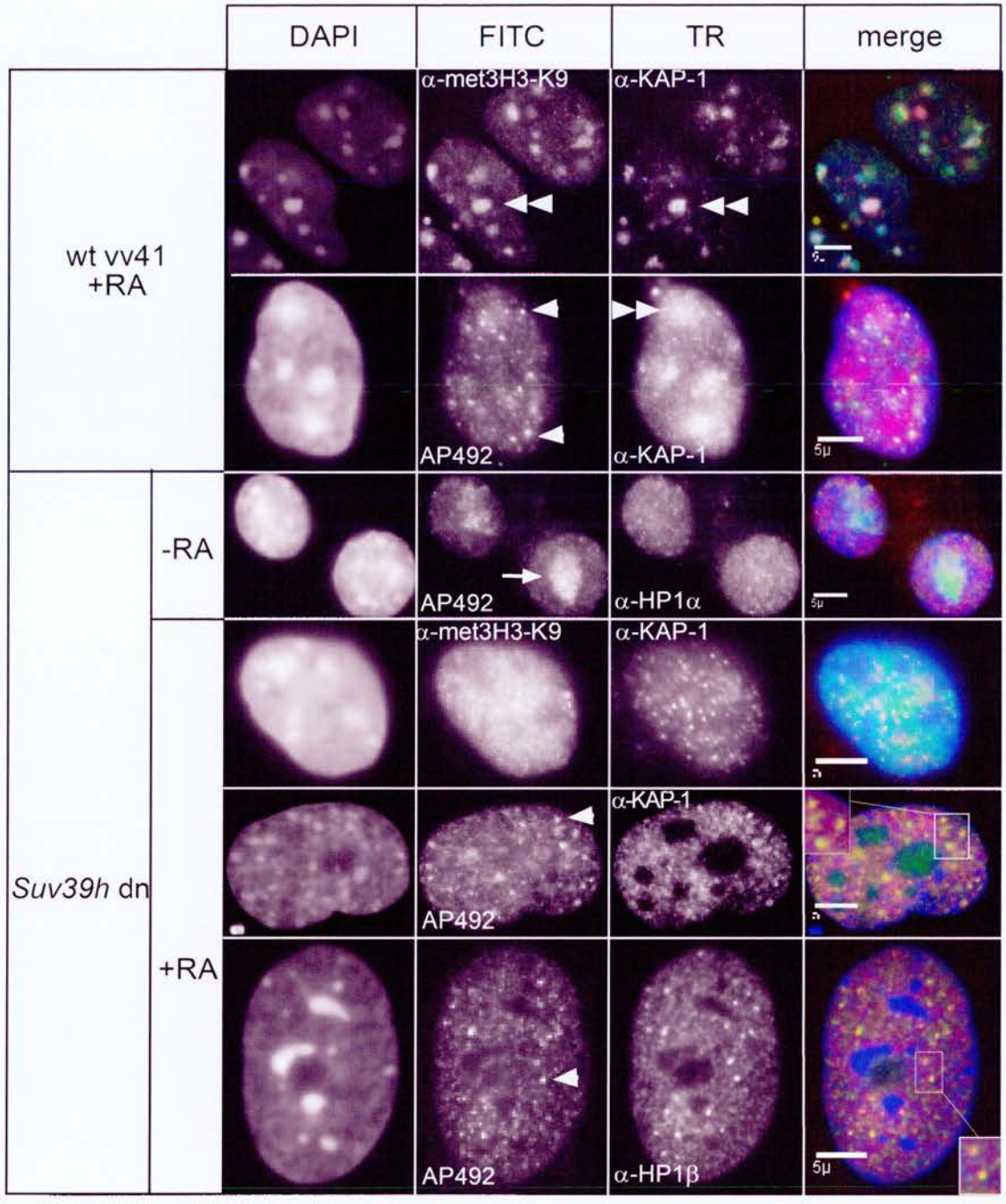


Figure 5.6 Analysis of sub-nuclear distribution of KAP-1 and KRAB-ZFP 492 in *Suv39h* dn cells

Immunofluorescence of undifferentiated (-RA) and differentiated (+RA) wt vv41 and *Suv39h* dn cells with combinations of either 1:100 AP492 (green in merge), α-met$_3$H3-K9 (red/green in merge), α-KAP-1 or α-HP1 antibodies (both red in merge). Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining, arrowheads indicate punctate foci and double arrowheads indicate heterochromatic staining. Bars = 5 μm.

for the formation of these foci. If these foci are 'silencing factories' then SETDB1 is likely to be the HMT responsible for histone methylation at these euchromatic sites.

I also analysed KAP-1 and AP492 localisation in *Dnmt3a/b*$^{-/-}$ undifferentiated cells and in contrast to *Suv39h* dn cells, KAP-1 localisation at heterochromatin was not impaired (Fig. 5.7). Therefore, methylated DNA at pericentromeric heterochromatin is not a pre-requisite for KAP-1 localisation. Ayyanathan *et al.* (2003) suggests that DNA methylation is a secondary response to KRAB-KAP-1 mediated repression and is involved in the heritability of genes targeted for silencing by KRAB-ZFPs. These results suggest that KAP-1 localisation to heterochromatin is not the result of a repressed gene being targeted for DNA methylation. Rather any DNA methylation at the target gene is a separate event to KAP-1 function at heterochromatin. I also attempted to differentiate the *Dnmt3a/b*$^{-/-}$ cell line with RA to observe whether DNA methylation is required for the formation of non-pericentromeric foci of 492 and KAP-1. However, after a couple of days, all cells had died and therefore these could not be studied. Recently, *Dnmt3a/b*$^{-/-}$ cells were found to be unable to differentiate by removal of LIF, but remained viable and they retained stem cell-like characteristics such as *Oct4* expression (Jackson *et al.*, 2004). Therefore *Dnmt3a/b*$^{-/-}$ cells appear to be sensitive to RA, and these experiments suggest that DNA methylation is crucial in the RA response pathways employed by ES cells.

5.6 Discussion

The results presented in this chapter suggest localised chromatin changes rather than a translocation of KRAB-ZFP target genes to pericentromeric heterochromatin for KRAB-KAP-1 mediated repression. I found that neither endogenous KRAB-ZFP 492 nor NT-2 localised to pericentromeric heterochromatin. Rather, I have identified foci of KRAB-ZFPs that form in differentiated cells. Intriguingly, as well as being present at pericentromeric heterochromatin in differentiated cells, KAP-1 and HP1s are also found associated with these foci. I have named these foci KRAB and KAP-1 associated foci (KAKA foci). Finally in *Suv39h* dn cells which lack H3-K9 trimethylation, KAP-1 and HP1 were delocalised from pericentromeric heterochromatin, but were still observed, along with KRAB-ZFP 492, in KAKA foci.

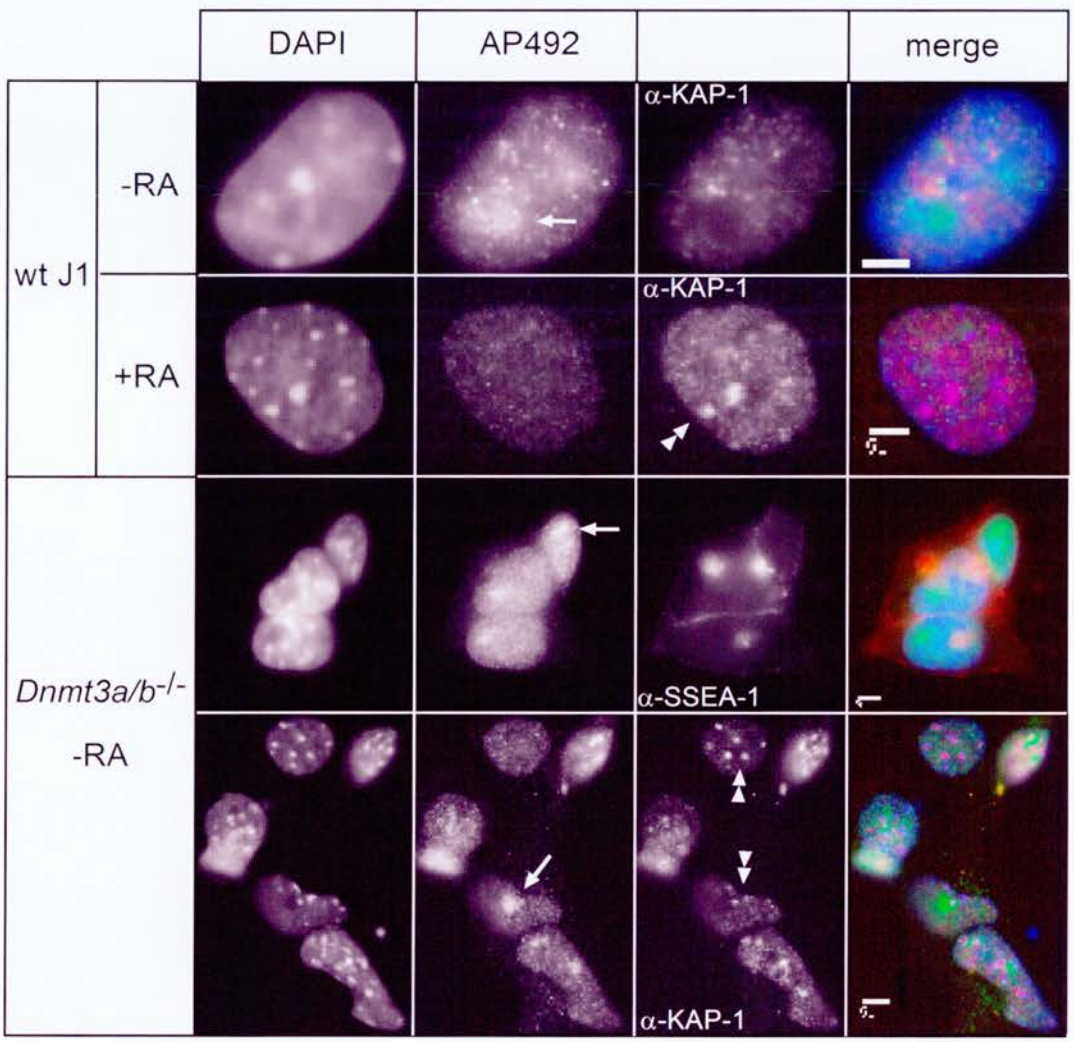


Figure 5.7 Analysis of sub-nuclear distribution of KAP-1 and KRAB-ZFP 492 in *Dnmt3a/b*$^{-/-}$ cells

Immunofluorescence of undifferentiated (-RA) and differentiated (+RA) wt J1 and *Dnmt3a/b*$^{-/-}$ cells with 1:100 AP492 (green in merge) and either α-KAP-1 or α-SSEA-1 antibodies (red in merge). Nuclei are counterstained with DAPI (blue in merge). Arrows indicate cells with nucleolar staining and double arrowheads indicate heterochromatic staining. Bars = μm .

I propose that the KAKA foci are sites of KRAB-KAP-1-HP1 mediated gene repression. As they still appear to form in *Suv39h* dn cells, this would occur via a Suv39h1/2 independent mechanism, even though Suv39h1 has previously been reported to have some euchromatic roles (Firestein *et al.*, 2000; Nielsen *et al.*, 2001a). In this way specific repression of KRAB-ZFP target genes would therefore be predicted to occur in euchromatic regions by construction of facultative heterochromatin during specific stages of proliferation or differentiation. As the SETDB1 HMTs has been found to interact with KAP-1, it is likely that in this system it cooperates with KAP-1 and HP1s to mediate repression of KRAB-ZFP target genes (Schultz *et al.*, 2002; Ayyanathen *et al.*, 2003). SETDB1 has previously been shown to have a euchromatic distribution in NIH 3T3 cells, but it would be interesting to investigate the localisation of SETDB1 (ESET) with respect to KAKA foci in differentiating cells (Schultz *et al.*, 2002). SETDB1 catalyses the conversion of histone H3-K9 di- to tri-methylation but does not appear to affect global H3-K9 tri-methylation or DNA methylation (Dodge *et al.*, 2004). Instead, the favoured model is that SETDB1 regulates gene expression at specific euchromatic sites (Schultz *et al.*, 2002; Dodge *et al.*, 2004). SETDB1 can methylate H3-K9 despite H3-K4 methylation, in stark contrast to Suv39h1, whose enzymatic ability is severely compromised by methylated H3-K4 (Schultz *et al.*, 2002). Therefore SETDB1 may function to repress KRAB-ZFP target genes in specially 'marked' regions of chromatin. It has also been suggested that SETDB1 may be instrumental in maintaining H3-K9 methylation at boundary elements, thus preventing the spread of transcriptionally active chromatin domains from juxta-posed regions enriched in H3-K4 methylation (Schultz *et al.*, 2002). In this way SETDB1 would be a good candidate for inducing facultative heterochromatic in KAKA foci. HP1 deposition at the *tk* promoter was totally dependent on H3-K9 methylation but optimal HP1 binding in an *in vitro* binding assay was also dependent on a functional chromoshadow domain for stabilisation of the H3-K9-HP1 interaction (Schultz *et al.*, 2002). The chromoshadow of HP1 interacts with the HP1BD of KAP-1 and would therefore be considered important for foci formation. Since KAP-1 exists as a homotrimer in solution and two HP1 proteins bind per KAP-1 monomer (Peng *et al.*, 2000; Lechner *et al.*, 2000), this might potentiate the propagation of facultative

heterochromatin at a target locus, but whether the chromodomains of a single HP1 molecule can bind to the methylated H3-K9 tails of a single or two adjacent nucleosomes remains to be answered.

KAKA foci in differentiated cells were observed with both AP492 and NT-2 antibodies. I have not been able to show if both antibodies occupy the same KAKA sites, as the antibodies were both raised in rabbit. I also attempted direct labelling of AP492 with fluorescently labelled IgG, but this was not successful. In order to answer this question I would need to raise an antibody in a different organism. It is not yet known whether KAKA foci are euchromatic regions specified by the presence of KAP-1-HP1 repression machinery or by the KRAB-ZFP target genes themselves. Sub-nuclear distribution of TIF1α suggests the presence of TIF1α docking sites where it functions as a chromatin docking protein to which liganded nuclear receptors bind (Remboutsika *et al.*, 1999). The KAKA foci characterised in this chapter could represent similar docking sites or 'silencing factories', which form during differentiation so that many KRAB-ZFPs can recruit their target genes for KAP-1-HP1 mediated silencing. The ATDC5 differentiation system in conjugation with the NT-2 antibody would be ideal for testing whether a target gene, for example, *col11a2*, is recruited to KAKA foci for silencing. However, due to both technical issues and limited amounts of NT-2 antibody, I was not able to do this critical experiment. Interestingly, like KRAB-ZFP 492, NT-2 was also found to be in or adjacent to the nucleolus in undifferentiated cells. This suggests either a nucleolar role for these proteins during cellular proliferation or that spatial organisation of KRAB-ZFPs within the nucleus may be important for regulating their activity.

If KRAB-ZFPs are not recruited to heterochromatin for silencing, then the enrichment of KAP-1 at heterochromatin in differentiated ES and ATDC5 cells suggests that it may have a function there other than KRAB-mediated repression. For example, heterochromatic KAP-1 localisation has been proposed as a KAP-1 storage site (Cammas *et al.*, 2002; Cammas *et al.*, 2004). In this regard accumulation of KAP-1 into regions of heterochromatin may control KAP-1 levels in the

nucleoplasm therefore creating different sub-nuclear compartments of KAP-1 which may exert different functions. In this respect, KAP-1 was reported to act as a co-activator and physically interact with both the glucocorticoid receptor (GR) and C/EBPβ to augment C/EBPβ transcriptional activity at the promoters of endogenous C/EBPβ-responsive genes during macrophage differentiation (Chang *et al.*, 1998; Rooney *et al.*, 2001). Whether KAP-1 has both co-activator and co-repressor functions in the same nuclei and whether different KAP-1 functions are separated *in vivo* by nuclear compartmentalisation remains to be determined.

Importantly, Cammas *et al.* (2004) reported KAP-1 localisation at heterochromatin to be critical for PE and VE differentiation. During differentiation of F9 cells, the KAP-1-HP1 interaction, and subsequent translocation to heterochromatin was required during a small window of time after the onset of RA treatment for complete differentiation into PE cells. However, my results show that in *Suv39h* dn cells, the translocation of KAP-1 to heterochromatin is abolished. This mis-localisation is certainly not critical to survival of cells or even the whole organism as 33% of knockout mice are viable (Peters *et al.*, 2001). Interestingly KAP-1 preferentially associates with regions of heterochromatin in the differentiated Sertoli and spermatogenic cell types (Weber *et al.*, 2002). *Suv39h* dn males display a complete spermatogenic failure at the transition between early to late spermatogenesis, and with delayed entry into meiotic prophase triggering apoptosis of spermatocytes during the mid- to late pachytene (Peters *et al.*, 2001). Therefore, KAP-1 mis-localisation in *Suv39h* dn cells may be related to the inability of spermatocytes to differentiate. Hence, KAP-1 localisation to heterochromatin could occur at crucial, decision-making stages of development. I would predict from localisation of the gene-trap fusion proteins, that any KRAB-box only splice forms of KRAB-ZFPs would localise with KAP-1. Also, I cannot rule out that a transient association of KRAB-ZFPs and their target loci to heterochromatin may also be important for their repression as suggested by Ayyanathen *et al.*, 2003. On this note, KRAB-ZFP 492 was not specifically excluded from heterochromatin in my studies.

Neither KAP-1 nor HP1 re-localised from pericentromeric heterochromatin with the formation of KAKA foci. However, the importance of HP1 localisation at euchromatic sites has recently been demonstrated. A dynamic re-location of HP1α from centromeric heterochromatin to the nucleoplasm was found to occur during the transition from proliferation to quiescence, in contrast to HP1β remaining at heterochromatin (Grigoryev *et al.*, 2004). This re-location coincides with the formation of facultative heterochromatin and transcriptional down-regulation, hence HP1α is predicted to assist in this global chromatin repression and facultative heterochromatin formation. The correct nuclear distribution of euchromatic HP1 proteins might be crucial for the various regulatory aspects of cell proliferation and differentiation. This has been demonstrated most recently in ICF patients who show aberrant HP1 signal of all 3 isoforms in the under-condensed 1qh or 16qh chromatin, which is thought to be the result of prolonged association of HP1 with satellite DNA as a consequence of DNA hypomethylation (Luciani *et al.*, 2005).

Finally the step-wise model of KAP-1 co-ordinated gene repression at euchromatic sites shares similarity with only one other co-repressor, Rb (retinoblastoma protein). The Rb protein co-ordinates similar activities in the repression of euchromatic E2F target genes in bringing the Suv39h1 enzyme to the promoters of E2F target genes where it can methylate K9 of histone H3 and provide a binding site for HP1 (Nielsen *et al.*, 2001a). The HDAC activity associated with Rb may be a preceding step to Suv39h1-mediated methylation since methylation cannot take place on an already acetylated lysine. Likewise, for KRAB-KAP-1 repression, the HDAC activity associated with the NuRD complex may precede SETDB1 methyltransferase activity allowing HP1 binding and subsequent heterochromatinisation of the promoter specified by a particular KRAB-ZFP (Schultz *et al.*, 2002). The visualisation of these foci in cells provides further evidence of promoter heterochromatinisation. The key question is whether these foci are associated with endogenous KRAB-ZFP target genes and this remains the key unanswered question in this field of research. A way to identify the physiological function and hence target genes of KRAB-ZFP 492 is to generate knockout mice and compare the gene expression profiles of wt and

homozygous mutant mice. The final results chapter of this thesis will describe the initial results of these studies.

Chapter 6

Establishment and characterisation of mutant KRAB-ZFP mice

6.1 Introduction

The effective trapping and mutating of genes with the *β-geo* cassette provided a way in which to study the KRAB-ZFP family that had not previously been utilised. I wanted to establish some mutagenic KRAB-ZFP mouse lines from gene-trapped cell lines as a first step in the initiation of their functional analysis. Since in mice, no KRAB-ZFP target genes were known at the start of my PhD, it was thought that the establishment of transgenic mouse lines, in addition to giving specific clues as to the spatial and temporal expression pattern of a particular KRAB-ZFP, would eventually prove to be useful for target gene assessment.

Only two mice have been reported with mutant KRAB-ZFP function (Casademunt *et al.*, 1999; Krebs *et al.*, 2003) although KRAB-ZFP genes are deleted in the embryonic lethal mutation t^{w18} (Crossley *et al.*, 1991; Shannon *et al.*, 1999). NRIF (neurotrophin receptor interacting factor) is a ubiquitously expressed KRAB-ZFP of the A+B KRAB box family (Casademunt *et al.*, 1999). It physically interacts with tumour suppressor protein, $p75^{NTR}$, which is a neurotrophin receptor of the TNF (tumour necrosis factor) receptor family of transmembrane proteins. Mice carrying a deletion of the $p75^{NTR}$-binding domain (pseudofinger region) and four of the five C2H2 ZFs of NRIF, cannot survive beyond E12 in a congenic BL6 strain. In the Sv129 background, mice are viable and healthy to adulthood. In the retinal cells of these mice, NRIF appears to play a role in $p75^{NTR}$ mediated apoptosis since retinal cells in mutant mice exhibit reduced cell death, a phenotype similar to that of $p75^{-/-}$ mice (Frade and Barde, 1999). More recently, Krebs *et al.* (2003) described the necessity of a pair of KRAB-ZFPs in sexually dimorphic liver gene expression during efforts to elucidate the gene responsible at the *Rsl* (regulator of sex limitation) locus. This locus confers repression of *sex-limited protein* (*Slp*) that is normally expressed in adult males only. Recessive *rsl* mutant alleles have arisen in populations of inbred mice and result in aberrant expression of *Slp* in the livers of female mice at 3 weeks of age and increased expression of *Slp* in males (Tullis *et al.*, 2003). In transgenic *rsl* mice, two KRAB-ZFP paralogues (*Rslcan-4* and *Rslcan–9*) have 94% protein similarity, and were consistently polymorphic in *rsl* strains (Krebs

et al., 2003). Each of these genes was subsequently found to restore the appropriate expression of a distinct sub-set of the male-specific liver genes. Therefore these proteins act together to perform the function of *Rsl* and the small number of amino acid differences between them is predicted to account for their different target selection (Krebs *et al.*, 2003).

From the work I have described in previous chapters, the gene-trap fusion proteins mis-localise in the nucleus compared to endogenous KRAB-ZFPs. Therefore I would predict the gene-trap lines to functionally disrupt the relevant KRAB-ZFP. They therefore provide an opportunity to expand the number of genes for which a mouse knockout is available. Also, as mapping KRAB-ZFP genes is difficult because of repetitive arrays, using gene-trap lines overcomes the challenge that may prevent these genes being knocked out by conventional means.

I have established transgenic mouse lines from the ES113 and ES492 cell lines because KRAB-ZFP 492 is well conserved between species and KRAB-ZFP 113 appears to be in a large rodent specific cluster (Chapter 3). I have performed an initial analysis on their spatial and temporal tissue expression and in doing so we are closer to analysing both the *in vivo* functions of KRAB-ZFPs 113 and 492 and investigating the consequences of their mis-localisation via the KAP-1 interaction.

6.2 Generation and establishment of ES113 and ES492 KRAB-ZFP transgenic lines

6.2.1 Generation of chimeric mice

KRAB-ZFP mutant mice were produced using gene-trapped ES cells from lines 113 and 492 for reasons discussed in section 3.6. Early passage ES cells (ES113p5 and ES492p6) were injected into C57BL/6 blastocyst recipients at 2.5 days post coitum (dpc) as described in section 2.6.2. Two chimeras (identified by a mottled coat colour), designated chimera 63 and 64 respectively were obtained from the ES492p6 blastocyst injections but only one chimera (chimera 65) was obtained from the

ES113p5 blastocyst injections. Each chimera was then backcrossed to the C57BL/6 inbred and the MF1 outbred strains (Table 6.1). Chimera 64 produced one offspring in a C57BL/6 cross, but fortunately transmitted the pGT gene-trap transgene. In crosses with C57BL/6, chimera 63 transmitted the gene-trapped chromosome to 50% of offspring, but produced few offspring with outbred MF1 mice of which none were *LacZ* positive. The ES113 chimera 65 transmitted the gene-trap locus to 60% of offspring when bred to C57BL/6, but these did not breed successfully with MF1 mice. The lack of MF1 offspring from chimera 64 and 65 is likely due to the fact that these chimeras had reduced breeding capability after the first few matings with C57BL/6, as opposed to mating ability being strain dependent.

Table 6.1 Genotype analyses of chimeric backcross progeny

Mouse line	*Chimera*	*Background*	*No. of progeny*	*No. of LacZ heterozygotes (+/LacZ)*
ES492p6	63	C57BL/6	4	2
		MF1	3	0
	64	C57BL/6	1	1
		MF1	0	0
ES113p5	65	C57BL/6	7	4
		MF1	0	0

6.2.2 Genotyping of transgenic mice for *LacZ* transgene expression

A genotyping PCR assay has previously been used in the lab to assay for the presence of the *LacZ* transgene, in mice generated from ES cells that carry the *β-geo* cassette (Heidi Sutherland Pers. Comm.). Heterozygous mice (het) (section 6.2.1) carrying both a *LacZ* negative, wt allele ('+') and a targeted disruption (*LacZ* positive) in either the 113 or 492 gene-trapped allele ('*LacZ*') were identified by PCR with primers specific for a 413 bp fragment of the *LacZ* transgene. The primers and amplification programme used are described in section 2.2.12 and can be found in Table 2.1. This method of heterozygous genotyping has been used throughout this

thesis for both the ES113 and ES492 transgenic lines to distinguish heterozygous mice ($113^{+/LacZ}$ or $492^{+/LacZ}$) from wt mice (113^{+} or 492^{+}) by visualisation of the 413 bp band after resolution of PCR products on a 2% agarose gel. Figure 6.1A shows the results from a typical genotyping PCR of offspring from a het x wt cross. A positive and negative control is performed for every batch of genotyping PCRs to confirm that the *LacZ* primers are working correctly in that reaction.

6.2.3 Establishment and propagation of ES113 and ES492 transgenic mouse lines

Chimeric offspring were subsequently backcrossed into both C57BL/6 and MF1 genetic backgrounds to produce both congenic inbred and outbred heterozygotes respectively for subsequent heterozygous intercrosses. Genotypes of whole litters were assayed by *LacZ* PCRs (section 6.2.2) and analysed for the predicted 1:1 Mendelian ratios of wt:heterozygote (if the heterozygotes are viable). Table 6.2 shows that the genotypic ratios for both transgenic lines backcrossed into the C57BL/6 background approximate to Mendelian ratios. Whilst this might also be true of the ES113p5 line backcrossed into MF1 (the slightly fewer than 50% heterozygotes (42%) might be correlated with the fewer number of mice analysed), the proportion of heterozygotes in the ES492p6 strain backcrossed to MF1 is only 36%. Therefore the fewer number of heterozygotes from the ES492p6 and possibly the ES113p5 transgenic lines when backcrossed into MF1 could initially suggest that the survival of heterozygotes may be compromised in the outbred line. Genotyping by PCR suggests a chance occurrence of false negative results, which must be considered when analysing the percentages of heterozygous offspring. There did not appear to be any correlation between the sex of progeny and their heterozygosity (data not shown) as would be expected of autosomal genes.

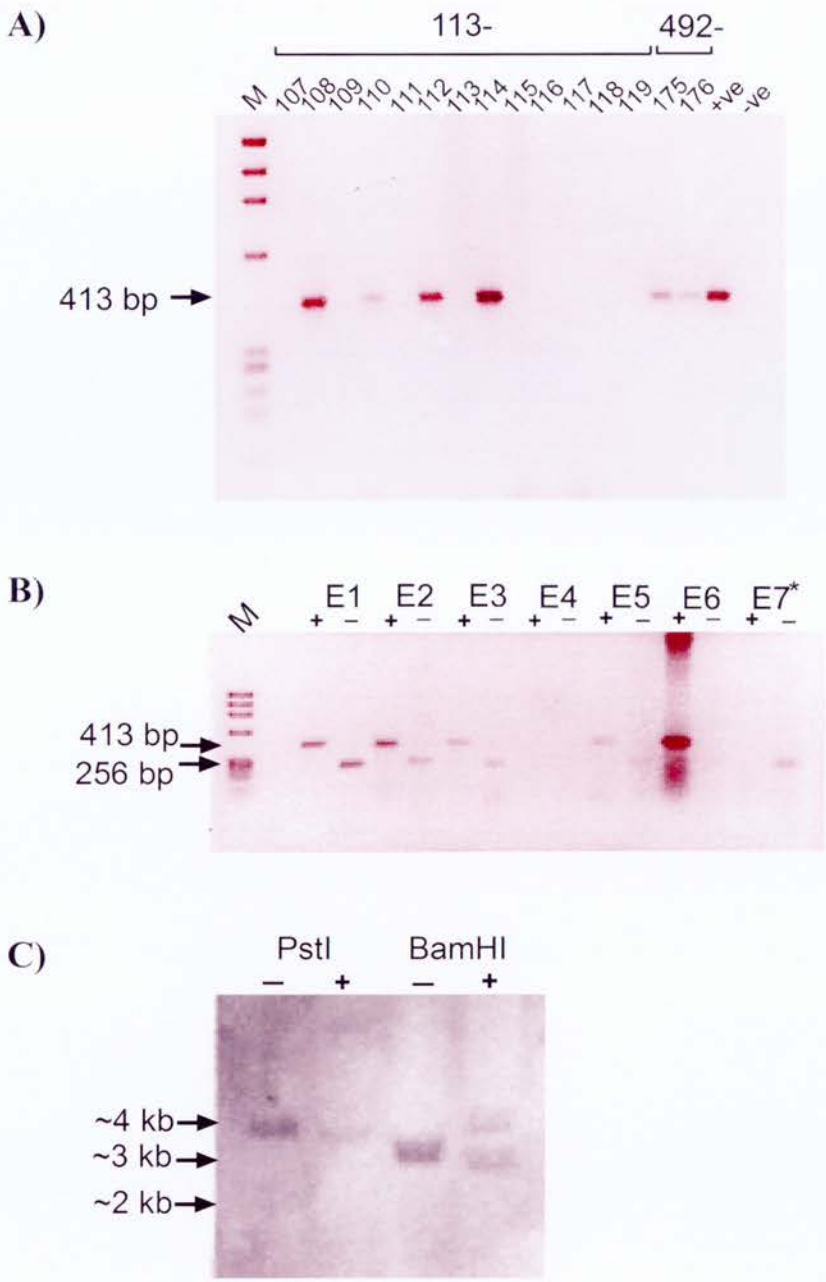


Figure 6.1 Genotyping of homozygous and heterozygous mice

A) PCR with primers specific for *LacZ* were performed on ~300 mg/ml mouse genomic DNA from either the ES113 or ES492 transgenic mouse lines. A 413 bp fragment can be seen (arrow) indicative of heterozygous mice (numbered 113-107 to -119 and 492-175 and -176) carrying the β-*geo* insertion. Positive (+ve) and negative (-ve) PCR controls are performed with DNA from confirmed transgenic mice carrying β-*geo* or with no DNA respectively. **B)** *LacZ* (+) and wt 492 (-) PCRs on 9.5 dpc embryos from a 492 het x het cross. Asterisks indicate wt mice. Embryo 6 (E6) was later confirmed as heterozygous (data not shown). **C)** Genotyping of KRAB-ZFP 492 wt (492-60) and heterozygous (492-70) mice by Southern blotting on 10 μg of genomic DNA digested with either PstI and BamHI restriction enzymes and with a [α-^{32}P] dCTP labelled PCR probe to the 492-linker region.

Table 6.2 Genotype analyses of $492^{+/LacZ}$ and $113^{+/LacZ}$ backcross progeny

Mouse line	*Background*	*Genotypic ratio*		*Unknown*	*Total*
		+/LacZ	*+*		
ES113p5	C57	44	42	11	97
	MF1	29	38	1	68
ES492p6	C57	35	40	10	85
	MF1	43	70	6	119

6.3 Production of homozygous mutant mice

Heterozygous mice from both lines were backcrossed at least three times into either the C57BL/6 or MF1 backgrounds and were subsequently mated for the production of mutant homozygotes (113^{LacZ} or 492^{LacZ}). Over 70 progeny per line were generated from these crosses and their genotypes analysed by *LacZ* PCR, and in the case of ES492p6 transgenic mice, by PCR of the undisrupted 492-linker region.

6.3.1 ES113

6.3.1.1 Genotyping of homozygous ES113 mice (113^{LacZ})

I was unable to decipher the gene trapped in the ES113 cell line by bioinformatics analyses (section 3.3.2). I originally chose to make transgenic ES113 mice on the basis that this gene-trap insertion was likely in a ZFP cluster with a view to addressing the question of KRAB-ZFP redundancy. Since the insertion is in a large ZFP cluster and the region is unassembled and unorientated, I cannot be sure that the gene-trap has only disrupted one gene. I therefore need more genomic sequence data to confirm the precise location of the gene-trap in this line. Using primers within the 5' RACE sequence and within either the SA region of *β-geo* or within *LacZ* itself, I endeavoured to PCR the intron sequence between the KRAB domain and where their gene-trap had inserted. All attempts were unsuccessful (data not shown). It is likely that the SA region of *β-geo* has been disrupted in some way during integration or that

the intron length is too great. A more comprehensive study of the genomic sequence surrounding the ES113 gene-trap integration is therefore needed and this study was not undertaken in this thesis. However, I continued to interbreed heterozygous mice to generate some DNA samples from offspring on the premise that homozygote detection might eventually be achieved.

6.3.1.2 Generation of homozygous mutant mice

The results for the ES113p5 line suggest a phenotype consistent with a mutant homozygous lethal mutation, which would give a ratio of 2:1 for *LacZ* positive offspring compared to 3:1 that would be expected if homozygotes (hom) survive. Out of 72 surviving progeny, 64% carry the *LacZ* transgene and 36% are wt (*LacZ* negative) (Table 6.3). In the C57BL/6 background, only 38 progeny out of 8 litters were obtained (~4 progeny per litter), 38% of which were found dead at various time points after birth. For this reason heterozygous crosses are now being continued in the MF1 background only, since litters with a greater number of progeny are consistently achieved in the outbred mice. The lack of progeny in the C57BL/6 background could be due to age dependent parental fertility problems as mice only appeared to breed once or twice with partners. However, it could be that some heterozygotes and homozygous embryos have survival problems in the C57BL/6 background. A similar scenario has been recorded for *nrif-/-* mice were embryonic death occurs at ~E12 in the C57BL/6 train but *nrif-/-* pups survive in the 129Sv inbred strain even after 10 backcrosses (Casademunt *et al.*, 1999). In ES113 het x het crosses (C57BL/6 background) I observed some re-adsorped embryos in litters at E12.5 and E16.5. Thus, it appears as though C57BL/6 heterozygous mice can interbreed, but perhaps there are also problems with the heterozygotes during embryonic development. Although the ES113 gene is in a large KRAB-ZFP cluster, heterozygous matings suggest its function is non-redundant. A greater number of progeny will need to be assessed in this line to conclude the phenotype of homozygous lethal.

Table 6.3 Genotype **analyses of 113$^{+/LacZ}$ intercross progeny**

Background	*Stage*	*Genotypic ratio*		*FD*[I]	*Un*[II]	*Re-adsorption*	*Total*[III]
		LacZ/?	+				
C57/BL6	E12.5	ND	ND			2/9	9
	E16.5	ND	ND			4/10	10
	Adult	5	7	14	12		38 (8L)
MF1	Adult	38	22	0	5		65 (5L)
Total (MF1+C57BL/6)	Adult	43	29	NA			72

[I] *FD*, found dead. The number of mice found dead between 1 to 23 days after birth. Their genotype was not determined.
[II] *Un*, unknown. The genotypes of these mice are unknown due to either the quality of the DNA and/or missing progeny whose genotypes were not determined.
[III] *L*, litters

6.3.2 ES492

6.3.2.1 Genotyping of homozygous ES492 mice (492^{+})

6.3.2.1.1 Genotyping by PCR

I wanted to develop a quick and simple genotyping assay to assess homozygous mutant 492 mice. I therefore designed PCR primers to the linker region of ES492 (Fig. 3.8, dashed arrows) that flank the gene-trap integration site. The primer sequences and PCR amplification programme are described in detail in section 2.2.12 and Table 2.1. By PCR and product resolution on a 2% agarose gel, these primers successfully amplified a 256 bp region of the wt 492 allele only as it was predicted that this primer set would be incapable of amplifying the pGT insert as well. The *LacZ* and wt 492 primers could not be combined in the same PCR without the generation of non-specific bands (data not shown). Therefore two sets of PCRs were performed on all progeny from a het x het cross (section 6.3.2.2). Wt mice would be predicted to carry two wt 492 alleles and thus only the 256 bp band is amplified by

PCR (Fig. 6.1B, asterisks). Heterozygotes carry both the wt allele (256 bp band) and the gene-trap allele (413 bp band). Homozygotes would be predicted to carry the gene-trap allele only (i.e. two copies of *LacZ*) and therefore only the 413 bp band would be seen on a gel.

6.3.2.1.2 Genotype confirmation by Southern Blotting

In order to confirm some of the homozygous and heterozygous mice as identified by PCR, I began the generation of a Southern blot assay. I made a PCR probe of the wt 492-linker region, i.e. the same region used for wt 492 PCRs (primers are described in Table 2.1) and labelled it with [α-^{32}P] dCTP as described in section 2.2.14.2. I decided to use the linker as a probe because this region is the most variable within the KRAB-ZFP family. The probe was used for Southern blotting of genomic DNA, which was digested with a variety of restriction enzymes beforehand (section 2.2.14). I wished to be able to detect a difference in banding pattern between wt homo- and heterozygous mice after their DNA had been digested with restriction enzymes. Initial results show that genomic DNA digestion with BamHI shows a clear difference in banding pattern between homozygous wt (492-60) and heterozygous mice (492-70) confirming previous identification by PCR (Fig. 6.1C). The size of the wt band in the BamHI digest was also confirmed bioinformatically (data not shown). PstI may also show a slight difference in banding pattern, although it is not as clear. It would therefore be expected that only the top band would be detected for a BamH1 digestion of DNA from a homozygote, however, I have not tested any homozygotes by Southern blotting.

6.3.2.2 Generation of homozygous ES492 mice

Similar to the ES113 transgenic line, the number of *LacZ* positive mice obtained from het x het crosses of ES492p6 mice suggests a phenotype consistent with a mutant homozygous lethal mutation i.e. close to 2:1. Out of 79 surviving progeny, 64% carry the *LacZ* transgene and 36% are wt (Table 6.4). Furthermore, using DNA obtained from 4-week adult mice, no homozygotes were identified by PCR. In the C57BL/6 background, 95 progeny out of 20 litters were obtained (~7-8 progeny per litter). ~13% of progeny were found dead shortly after birth and of these mice, the

majority were genotyped as heterozygotes, with 1 possible but as yet unconfirmed homozygote (Table 6.4). Timed C57BL/6 heterozygous matings were set up to collect embryos at ~ E8.5-9.5. The genotypes of the parents of these matings were confirmed as heterozygous by PCR but the mice were also used for backcrossing, with progeny genotypes reflecting the expected mendelian ratios, as further proof of their heterozygosity. Out of 29 progeny, the wt band (256 bp) for only 1 embryo could not be amplified by PCR. The percentage of heterozygous and wt mice is 41% and 44% respectively, although the DNA samples and therefore genotypes of 4 mice were not obtained. Although these numbers would suggest that some lethality is occurring in $492^{+/LacZ}$ mice, in the C57BL/6 background, the results of adult mice obtained for het x wt (Table 6.2) and het x het (Table 6.4) crosses do not corroborate this. More embryos at this time point are required to draw further conclusions from this data. Heterozygous matings with at least N_6 (6^{th} generation) backcrossed C57BL/6 mice yielded less progeny and in most cases of timed heterozygous matings, mice did not appear pregnant. The 492 mutation may affect fertility in the heterozygous state, but lack of progeny may also be independent of the ES492 gene-trap because backcrossing into C57BL/6 is known to result in impaired breeding. Therefore, I performed timed heterozygous matings in the MF1 background. At E8.5, there were no re-adsorped embryos, but in a litter at E13.5, three re-adsorped embryos were obtained. A greater number of progeny will need to be assessed in this line in order to conclusively describe this mutant phenotype as homozygous lethal and more timed matings are required to assess the time at which the lethality is occurring.

The lack of homozygous mice and the ratio of heterozygous to wt mice in this line suggest the function of KRAB-ZFP 492 is vital and non-redundant for survival. Moreover, the phenotype of homozygous 492 mice is likely to be linked to the mis-localisation observed with this protein due to deletion of its ZF domain.

Table 6.4 Genotype analyses of $492^{+/LacZ}$ intercross progeny

Background	*Stage*	*Genotypic ratio*[I]			*FD*[II]	*Un*[III]	*Re-adsorption*	*Total*[IV]
		Lac Z	*+/Lac Z*	*+*				
C57/BL6	E8.5-9.5	1?	12	13		4	0	29 (3L)
	Adult	**0**	**41**	**24**	**13 ($6^{+/LacZ}$, 1^{LacZ}?)**	**17**		**95(20L)**
MF1	E8.5	ND	ND	ND			0	13
	E13.5	ND	ND	ND			3	15
	Adult	0	10	4	0			14 (2L)
Total (MF1/ C57BL/6)	Adult	0	51	28	NA			79

[I] Question mark indicates mice whose wt band is repeatedly not amplified by PCR
[II] *FD*, mice found dead shortly after birth. The genotype was not determined for 6 of these mice. Question mark indicates mice whose wt band is repeatedly not amplified by PCR
[III] *Un*, unknown. The genotypes of these mice are unknown due to either the quality of the DNA and/or missing progeny whom had not been genotyped.
[IV] *L*, litters

6.4 Expression analysis of KRAB-ZFPs 113 and 492

The tissue specific expression of many KRAB-ZFPs has been studied mainly by northern Blot analysis (Bellefroid *et al.*, 1993; Lee *et al.*, 1997; Mark *et al.*, 1999; Kim *et al.*, 2000). Although these studies do not necessarily detect the presence of translated KRAB-ZFP they do potentially show the many different tissue specific and temporal patterns of protein expression during differentiation, ranging from very specific expression in various types of cells (Shannon *et al.*, 1999; Kim *et al.*, 2000; Mark *et al.*, 2001; Kim *et al.*, 2001; Looman et al., 2003) to otherwise ubiquitous expression patterns (Casademunt *et al.*, 1999; Dreyer *et al.*, 1999; Mark et al., 2001; Kim *et al.*, 2001; Dai *et al.*, 2003; Gou *et al.*, 2003). In fact, their broad and non-descriptive role as regulators of differentiation and development stems from the vast number of mRNA analyses and protein expression studies. Expression of some KRAB-ZFPs are controlled in a cell cycle or differentiation specific manner (Jheon

et al., 2001; Rue *et al.*, 2001; Jheon *et al.*, 2002; Zhou *et al.*, 2002; Luo *et al.*, 2002) whereby expression may remain in the same tissue but with changing levels throughout different stages of differentiation or development (Luo *et al.*, 2002; Tanaka *et al.*, 2002; Zhou *et al.*, 2002). Expression may vary between different tissues during development (Tanaka *et al.*, 2002; Zhou *et al.*, 2002).

I have generated expression data for the 113 and 492 KRAB-ZFPs, which will be essential for the future study of these proteins. Although at the sub-cellular level, their distribution is mis-localised, their temporal expression is under control of the endogenous protein as the same promoter and upstream regulatory elements of the endogenous gene are utilised by the gene-trap vector. The following expression studies are a result of het x wt matings in the C57BL/6 background.

6.4.1 Wholemount X-Gal staining of embryos and their analysis by paraffin wax sectioning

Both ES113 and ES492 genes were trapped in ES cells therefore I decided to examine their gene expression patterns at various stages of embryonic development. Firstly embryos were harvested from their mothers at various time points and freed from their extra-embryonic membranes and yolk sac using scissors and forceps (section 2.6.3). They were rinsed in PBS before fixation in either 4% pFa or gluteraldehyde fixative for 1 hr (section 2.8.4.1) and then staining o/n in X-Gal. For each litter, ~50% of embryos stained blue (indicative of X-Gal being metabolised by β-galactosidase) as expected from a het x wt cross. The wholemount expression patterns for each transgenic line at each stage of embryonic development analysed are described in sections 6.4.1.1 (ES113) and 6.4.1.2 (ES492) respectively. I then analysed every wholemount staining pattern further by processing embryos at each developmental stage in paraffin wax (section 2.8.5.1). Wax embedded embryos were subsequently sectioned using a microtome followed by dewaxing and staining in nuclear fast red (nfr), which stains cell nuclei (section 2.8.5.2). The results of these analyses are described concomitantly with the wholemount expression data below.

6.4.1.1 Tissue specific expression of KRAB-ZFP 113 during embryonic development

Figure 6.2 illustrates the tissue specific expression of the KRAB-ZFP 113-fusion protein at 8.5 and 10.5 dpc. During the early stages of embryonic development (E8.5-E10.5), KRAB-ZFP 113 is strongly expressed in the nuclei of a population of myocardial cells (Fig. 6.2B, B'+D') but is absent from cells of the pericardial wall and endocardial cushion (Fig6.2B, E+D'), which is associated with the wall of the atrio-ventricular canal. Expression can also be seen in the endothelial lining of the dorsal aorta (Fig6.2B, D+D'). Interestingly, cells in the location of the thymus primordium also stain intensely blue at E8.5 and E10.5 (Fig 6.2B, A+C). By 12 dpc, *LacZ* expression is found in the skin and appears more ubiquitous.

6.4.1.2 Tissue specific expression of KRAB-ZFP 492 in the during embryonic development

Wholemount X-Gal staining showed that KRAB-ZFP 492-fusion protein is strongly expressed from at least 8.5 dpc (data not shown). At 10.5 dpc, strong expression is found in the cephalic region (Fig. 6.3A). This region is involved in differentiation of the branchial arches and components of the face and nose and expression is observed in the eye (Fig. 6.3A) and nose (Fig. 6.3B, B) at all stages of development presented. Expression is also clearly seen in the hillocks on either side of the first branchial cleft, which are associated with the formation of the external ear (Fig. 6.3A, E11.5 and E15.5). From 10.5 dpc onwards, expression is seen in the limb buds (Fig. 6.3A and 6.3B, C) and in a distinct banding pattern that reflects the joints of the limb (Fig. 6.3A, arrows). At 10.5 dpc, clear expression of *LacZ* is observed in parts of the brain (Fig. 6.4B; D, E+F) and expression here continues to at least 15.5 dpc (Fig. 6.3A). Sectioning of 10.5 dpc embryos, gives a clear indication of the tissue specificity of KRAB-ZFP 492. X-Gal stained many tissues of the developing gut and noticeably the pericardial wall and endocardial cushions (Fig. 6.3B, A+C), which coincidently, were not stained in $113^{+/LacZ}$ embryos of the same age. A strong and specific expression pattern was observed in the ventral horn of the neural tube and in the dorsal root ganglia (Fig. 6.2A, E13.5 and Fig. 6.3B, F). The afferent (sensory)

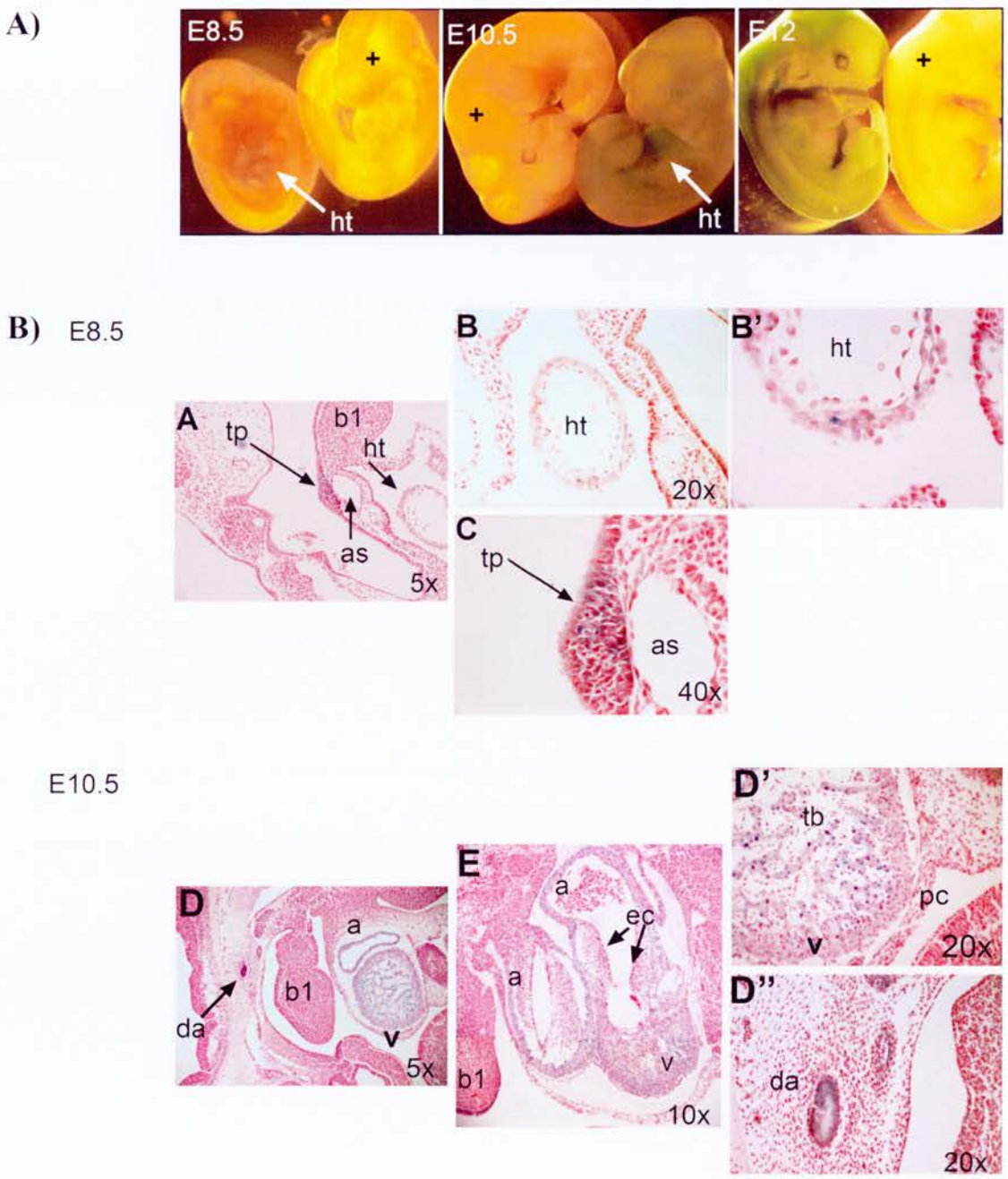


Figure 6.2 Tissue specific expression of KRAB-ZFP 113

A) Wholemount X-Gal staining (blue) of $113^{+/LacZ}$ embryos expressing *LacZ*. The respective control embryos are shown for comparison (+/+). **B)** *LacZ* expression in saggital sections of $113^{+/LacZ}$ embryos at 8.5 dpc(A-C) and 10.5 dpc (D-E). Expression can be seen in the myocardium of the common ventricular and atrial chambers of the heart (B,B',D,D',E), in the endothelial lining of the dorsal aorta (D,D'') and in the primitive thyroid (A, C). *a*, common atrial chamber of the heart; *as*, aortic sac (outflow tract from primitive heart); *b1*, mandibular component of first branchial arch; *da*, dorsal aorta; *ec*, endocardial cusion; *ht*, primitive heart; *pc*, pericardial wall; *tp*, thymus primordium; *tb*, trabeculated wall of common ventricular chamber of the heart; *v*, common venticular chamber. 5x, 10x, 20x and 40x magnifications.

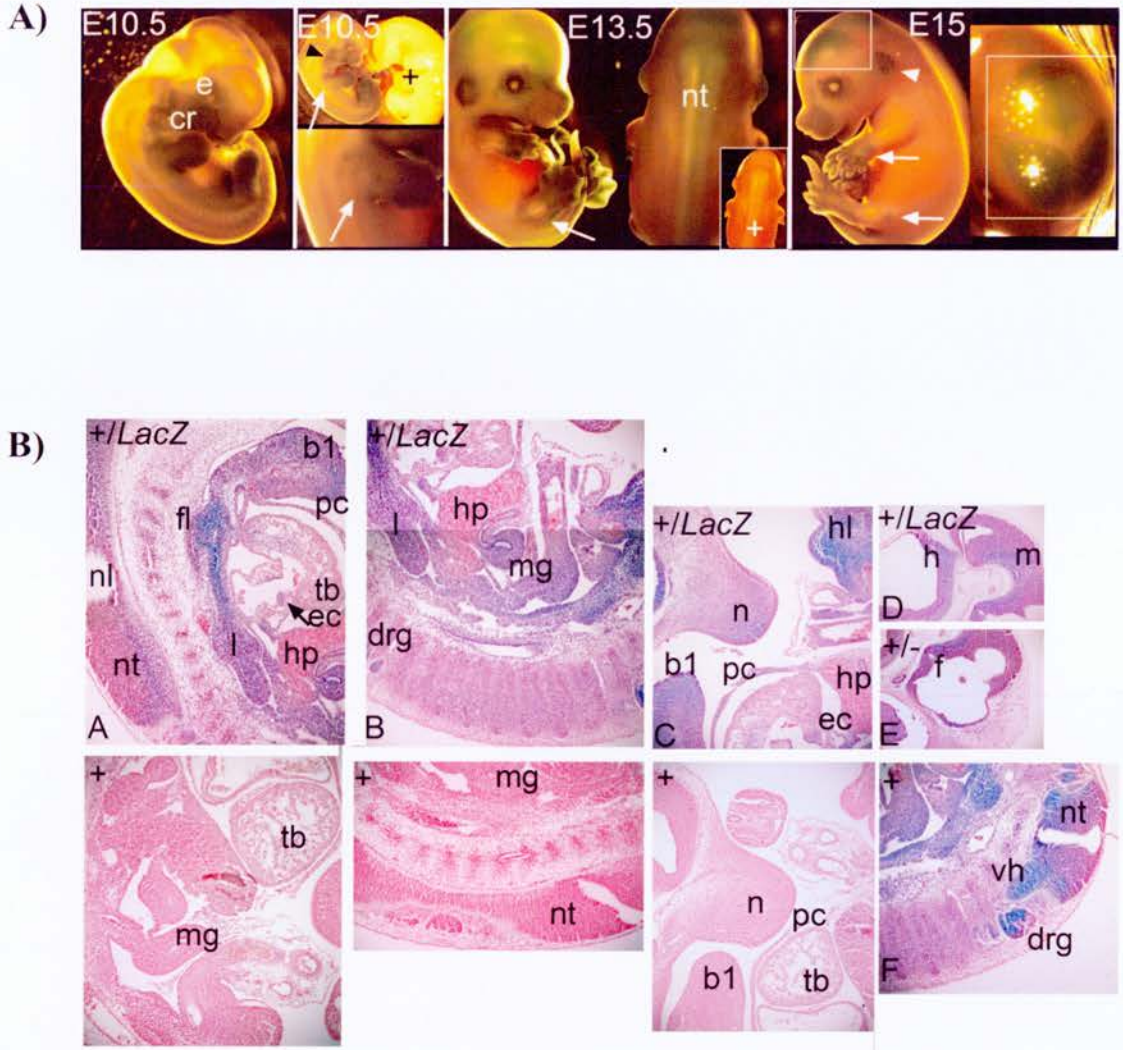


Figure 6.3 Tissue specific expression of KRAB-ZFP 492

A) Wholemount X-Gal staining (blue) of $492^{+/LacZ}$ embryos expressing *LacZ* . Control embryos (+) at E10.5 and E13.5 are shown for comparison. External expression is seen in limb joints (arrows) and in the ear (arrowheads). Strong expression is also found in the cephalic region (*cr*, branchial arches and fonto-nasal derivatives). **B)** *LacZ* expression in saggital sections of $492^{+/LacZ}$ and 492^{+} embryos at 10.5 dpc (A-F). Expression can be seen in various embryonic tissues derived fom all three germ layers. Note strong expression in the spinal cord (ectoderm) (A+F), inner epithelial linings of the gut (endoderm) (A+B) and the mesodermal derived branchial arches (A+B) and limb buds (B). *b1*, mandibular component of first branchial arch; *e*, eye; *ec*, endocardial cusion; *f*, neuroepithelium of forebrain; *fl*, fore limb bud; *h*, neuroepithelium of hindbrain; *hlb*; hind limb bud; *hp*, hepatic primordium (embryonic liver); *l*, lung bud; *m*, neuroepithelium of midbrain; *mg*, midgut; *n*, nose; *nl*, neural luman; *nt*, neural tube; *pc*, pericardial wall; *tb*, trabeculated wall of common ventricular chamber; *vh*, ventral horn. 5x and 10x magnifications.

neurones have their cell bodies in the dorsal root ganglia and carry information to the central nervous system. The efferent motor neurones have their cell bodies in the ventral (motor) horn and carry information to the muscles (Rhoades and Pflanzer, 1996), suggesting that KRAB-ZFP 492 may have a prominent role in the developing spinal cord. Nervous tissue including the brain is derived from ectoderm, cartilaginous tissue (and bone) is derived from mesoderm and linings of the gut, respiratory, and urogenital tracts are derived from endoderm (Kaufman, 1992). Therefore, KRAB-ZFP 492 is clearly expressed in all three germ layers deriving from the epiblast. It will be interesting in the future to identify the earliest time-point at which KRAB-ZFP 492 is expressed since no degenerating embryos were found at 8.5 dpc in progeny from het x het crosses.

6.4.2 KRAB-ZFP expression in adult mice

6.4.2.1 Wholemount X-Gal staining of adult tissues

In vivo expression of KRAB-ZFPs 113 and 492 was analysed in adult tissues by wholemount X-Gal staining of dissected organs (section 2.8.4) from confirmed heterozygous mice at least 4 weeks of age (Fig. 6.4A). Endogenous X-Gal staining was observed in some wt tissues; for example, in the bronchi of the lungs (probably bacteria), the medulla of the kidney and in the ovaries, uterus and testes. Of the tissues tested, KRAB-ZFP 113 only showed a comparable difference from wt tissues in the spleen and thymus. No expression was observed in the heart unlike during early embryonic development (E8.5-E10.5). Table 6.5 shows a comparison between the adult tissues stained in this analysis alongside the respective embryonic staining patterns. Although a splenic primordium can be observed at E8.5 and beyond, I did not observe X-Gal staining in the respective region at E8.5-E10.5. There was also a noticeable difference in X-Gal staining pattern between the thymus of $492^{+/LacZ}$ compared to wt mice. However, the X-Gal staining pattern appears different again between $492^{+/LacZ}$ and $113^{+/LacZ}$ mice (Fig. 6.4A, comparison is illustrated by arrows). The thymus is composed of many irregularly shaped lobules separated by connective tissue septae. Each lobule has an external cortex where T-cells become educated

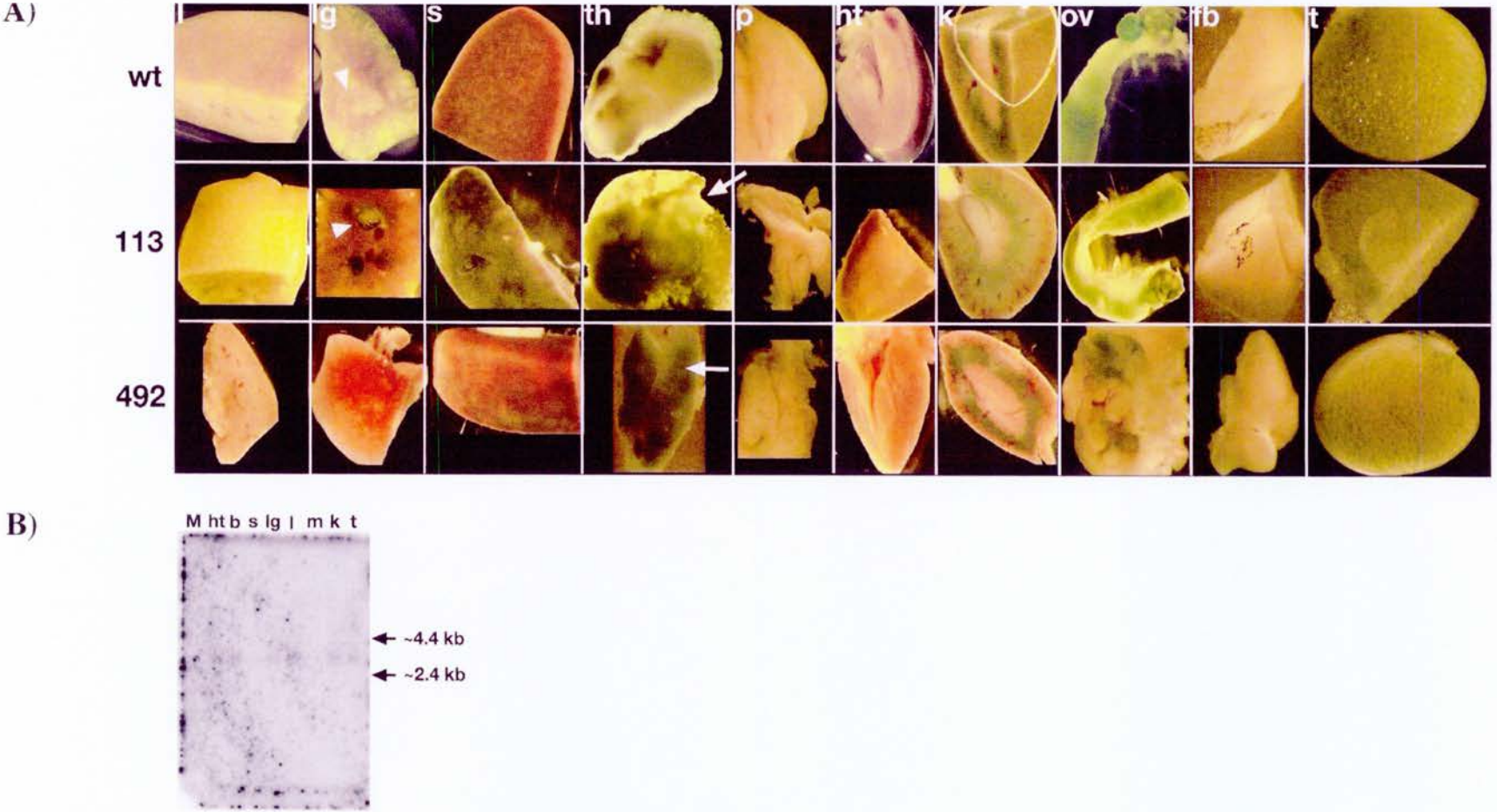


Figure 6.4 KRAB-ZFP 113 and 492 expression in adult tissues

A) Wholemount X-gal staining of adult tissues dissected from ES113 and ES492 heterozygous mice. Arrows indicate a difference in staining pattern of the thymus between KRAB-ZFPs 113 and 492. Endogenous staining in the bronchi of the lungs is shown by arrowheads. **B)** A Multiple Tissue Northern (MTN) Blot (BD Biosciences) of selected mouse tissues hybridised with [α-^{32}P] dCTP 492-linker probe. A single mRNA transcript is indicated (arrow) amongst a high level of background. *M*, marker; *b*, brain; *fb*, forebrain; *ht*, heart; *k*, kidney; *l*, liver; *lg*, lung; *m*, muscle; *p*, pancreas; *s*, spleen; *t*, testis; *th*, thymus; *ov*, ovaries and uterus.

(competent) and an inner medulla where mature thymocytes reside (Rhodes and Phlanzer, 1996). In the future it will be interesting to clarify the X-Gal staining pattern in both transgenic lines, which will give a more precise indication of their function.

Table 6.5 Comparison of KRAB-ZFP expression in embryonic and adult tissues

Tissue	*113*		*492*	
	embryo (E8.5-E10.5)	*adult*	*Embryo (E10.5)* [1]	*adult*
liver	-	-	-	-
lung	-	-	√	-
spleen	-	√	un	-
thymus	√	√	un	√
pancreas	-	-	un	-
heart	√	-	√	-
kidneys	-	-	√	-
ovaries	-	-	un	-
uterus	-	-	un	-
brain	-	-	√	-
testes	-	-	un	-

[1] *Un*, unknown. X-Gal staining in these primitive embryonic organs has not been confirmed.

6.4.2.2 Preliminary northern blot analysis of KRAB-ZFP 492 mRNA

To identify the number of KRAB-ZFP 492 mRNA transcripts in selected adult tissues and to lend support to X-Gal expression data I performed a northern blot on RNA samples from selected adult tissues (section 2.3.3). I used the same [α-^{32}P] dCTP labeled PCR probe to the linker region of 492 as I used for Southern blotting (section 6.3.2.1.2). In initial experiments with this probe, I could detect a faint mRNA transcript of approximately 3 kb (Fig. 6.4B), although it gave high background and so was not very clear. This suggests that in some adult tissues (for

example, liver, kidney and testes, there is only one splice-form of KRAB-ZFP 492 that contains the linker region. By western blot I could detect two possible protein species of very similar size in whole cell extracts (Fig. 4.10C). Therefore, it is possible that these two bands are due to post-translational modifications of the same protein isoform, or it is possible that I have not detected every mRNA splice-form since a probe to the linker region was used. The number of mRNA transcripts during embryogenesis remains to be determined.

There is some discrepancy between *in vivo LacZ* expression data and tissues transcribing KRAB-ZFP 492 mRNA as determined by northern blotting. For example, a possible mRNA band is observed in the liver by northern blotting, however in X-Gal stained tissue, the level of X-Gal staining is comparable to wt levels. However, due to technical reasons the latter experiment is confounded by high levels of background on the northern blot, and these experiments must be repeated to clarify these preliminary results.

6.5 Discussion

The gene encoding KRAB-ZFP 492 is in a non-KRAB-ZFP clustered region of the genome and has homologues in rat and human (chapter 3). As the gene-trap fusion protein mis-localises and as cells transfected with GFP-tagged 492 do not survive (chapter 4), my prediction was that this null-function KRAB-ZFP line would give an important phenotype and this appears to be the case. However, since ES113 likely resides in a KRAB-ZFP cluster, it will be very interesting to determine whether the knockout of this gene really causes lethality. *Rslcan-4* and *–9* of the *rsl* locus are amidst a 22 KRAB-ZFP cluster (Krebs *et al.*, 2003). Whilst it has been speculated that many members of KRAB-ZFP clusters may be redundant, mutations in *Rslcan-4* and *–9* expose the importance of individual but highly similar cluster members (Krebs *et al.*, 2003). My data, as a result of establishing mouse lines with disrupted KRAB-ZFP genes, also appears to show the importance of two more KRAB-ZFPs and in the future it may be possible to use the ES113 line to address the question of redundancy between the many members of this KRAB-ZFP cluster.

In a C57BL/6 background, the KRAB-ZFP 113 and 492 mutation appears to be embryonic lethal either before E12.5 or between E8.5 and E13.5 respectively but further work is needed to clarify the exact stage of embryonic development when these KRAB-ZFPs are required. For KRAB-ZFP 492, the 3 re-adsorped embryos at E13.5 appeared to be at different stages of re-adsorption (data not shown). One of these embryos has since been confirmed as homozygous mutant, whilst the remaining two are heterozygotes (data not shown).

Expression data suggests that the 113 and 492 KRAB-ZFPs are involved in differentiation of specific tissues during embryogenesis. Coincidently, both proteins appear to be expressed in the adult thymus. There are a number of different cell types in the thymus and it will be interesting to identify which cells express these proteins and if they are expressed in the same cell type. A number of KRAB-ZFPs are expressed in either the thymus alone or in the spleen and thymus (Bellefroid *et al.*, 1993; Rue *et al.*, 2001; Hennemann *et al.*, 2003). In particular, a whole cluster of human KRAB-ZFPs (the ZNF91 cluster on HSA19) have been shown to have a significantly higher level of expression in human T-cells but no equivalent cluster in mouse has been identified (Bellefroid *et al.*, 1993). Northern blot analyses has indicated that ZNF382 is broadly expressed in embryonic tissues, but similar to KRAB-ZFP 492, is tissue restricted in adults where it is expressed in heart (Luo *et al.*, 2002). However, KRAB-ZFP 113 and 492 may also be expressed in many other tissues on which I have not performed wholemount staining. The weak level of X-Gal staining in the skin of $113^{+/-}$ mice at 12 (Fig. 6.2A) and 13.5 dpc (data not shown) may be indicative of its expression their in adult mice and these possibilities should be born in mind for future experiments.

This expression data will be vital for pinpointing the organs/tissues to focus on, with respect to finding a cause for the phenotypes observed, and for future target gene identification. It will be important to find target genes and to determine whether their nuclear organisation is important for KRAB-ZFP repression, i.e. at KAKA foci, heterochromatin or elsewhere. Cloning of ChIPed sequences with AP492 antibody would be one way to obtain target sequences. Comparison of micro-array gene

expression data from homozygous wt and homozygous mutant tissues will also indicate genes with differential expression that may be directly regulated by either KRAB-ZFPs 113 or 492 and this data could enhance ChIP experiments. For micro-arrays, it is important to choose a tissue were these KRAB-ZFPs are being expressed. In the case of KRAB-ZFP 492, this could be any number of embryonic tissues, but the choice of tissue for 113 is more restricted. Expression data will also be useful for future yeast two hybrid experiments were the choice of cDNA library may be important for the discovery of protein interactors

To conclude, the results presented in this chapter show that KRAB-ZFPs 113 and 492 may be vital for embryogenesis. Moreover, mice with a disrupted KRAB-ZFP 492 gene exhibit a mis-localisation of KRAB-ZFP 492 upon differentiation due to interaction with KAP-1 at pericentromeric heterochromatin. I predict that such an interaction may cause aberrant regulation of KRAB-ZFP 492 target genes. In this case, the ensuing mis-regulation of genes may therefore be detrimental to the development of homozygous mutant mice, but more experiments are required to definitively prove KRAB-ZFP 113 and 492 are vital for embryogenesis.

Chapter 7

Discussion

7.1 Project development enabled by emerging resources and on-going research

In this thesis I have identified and characterised novel members of the largest transcription factor family in the mammalian genome, the KRAB-ZFP family (Chapter 3). Initially, my choices in regard to the specific KRAB-ZFPs I proposed to study were restricted by lack of complete sequence characterisation of most of the KRAB-containing genes. During the course of my PhD, I was able to characterise most of the remaining genes due to vast improvements in the annotation and sequence quality of mouse genome databases. This enabled me to draw parallels between the trapped genes and broaden my understanding of their genomic structure.

As a way of assessing the biological function of KRAB-ZFPs in ES cells, I studied their sub-cellular distribution before and after cell differentiation (Chapter 4). During my thesis the re-localisation of KAP-1 to pericentromeric heterochromatin upon cellular differentiation was reported (Cammas *et al.*, 2002). However there was no analysis of KRAB-ZFPs in this paper. I decided to shift the focus of my localisation studies to the analysis of endogenous KRAB-ZFP distribution compared with gene-trap distribution. My data, in conjunction with KRAB-ZFP epitope-tagging experiments and a thorough review of the past and emerging literature, provided me with new insights into the relationships between function of these proteins and their sub-nuclear distribution. Contemporary with the development of my project, it became clear that the concerted role of KAP-1 and HP1 in gene-repression may occur at euchromatic sites (Schultz *et al.*, 2002; Ayyanathan *et al.*, 2003). I therefore proposed to explore the distribution of KRAB-ZFPs in this context. In Chapter 5, I addressed how endogenous KRAB-ZFP distribution fits in with the current model of KRAB-KAP-1 mediated repression (Chapter 5).

To date, my thesis demonstrates the first thorough study on the sub-nuclear distribution of a KRAB-ZFP. My data, confirms a more general euchromatic role for KRAB-ZFPs whilst proposing an explanation for the presence of KAP-1 at

heterochromatin and the enigma that is pericentromeric localisation of KRAB-ZFPs. Additionally, I propose a novel role for the nucleolus in KRAB-ZFP sequestration.

In the latter year of my studies (2004), a second KRAB-ZFP knockout mouse was published, exposing for the first time how essential individual members of this protein family are during development (Krebs *et al.*, 2003). In Chapter 6, I describe the generation of mutant KRAB-ZFP mouse lines to further study the biological functions of novel KRAB-ZFPs and to investigate the importance of KRAB-ZFPs during development. These studies show that at least two other KRAB-ZFPs are critical for embryonic development and in the future it may be possible to use these results to address the question of redundancy between the many members of the KRAB-ZFP family.

7.2 Identification and characterisation of novel KRAB-ZFPs

A number of KRAB-containing proteins, originally isolated in a previous gene-trap screen, were identified by 5' RACE, FISH mapping and by sequence comparisons with genomic databases (Chapter 3). Five of the trapped genes were novel KRAB-ZFPs, two of which I chose for further study (KRAB-ZFP 113 and 492).

The distribution and genomic structure of the KRAB-ZFPs identified in this study enforce previous observations (Lichter *et al.*, 1992; Eichler *et al.*, 1998). At least 4 of these genes (ESKN205, ES261, ES492 and ES510) are found at breaks in synteny between the mouse and human genomes signifying the independent expansion of KRAB-ZFP clusters at such boundaries in different species (Shannon *et al.*, 2003). Some of the genes identified were found to have related sequences in either the rat genome alone (ES9, ES113, ESKN205) or both human and rat genomes (ES261, ES492) but further sequence analysis is required to confirm their orthology.

I identified three genes that potentially encode proteins containing the KRAB domain only (by virtue of their similarity to KRAB-only encoding cDNAs) and that lack any

ZF domain. Such proteins have been described previously as functionally important (Li *et al.*, 2003; Oh *et al.*, 2005). Alternative spicing of ZnF197 results in a KRAB-only protein named VHLaK. VHLaK mediates pVHL binding and subsequent repression of HIF-1α transactivation (Li *et al.*, 2003). Furthermore, VHLaK, pVHL and KAP-1 were found in a single complex suggesting that KAP-1 may participate in this repression. KRAB domain removal by alternative splicing may contribute to other biological pathways by redirecting protein interactions. The impact of alternative splicing on protein interactions in the human genome has recently been assessed (Resch *et al.*, 2004). Bioinformatics analysis of the proteome showed that a major impact of alternative splicing is removal of protein-protein interaction domains that mediate key links in protein interaction networks. Furthermore, the KRAB domain was highlighted as being selectively removed by alternative splicing at a much higher frequency than average, although C2H2 ZF domains were not listed among the 50 most common domains to be removed.

7.3 The implications of endogenous KRAB-ZFP localisation

In Chapter 4, I investigated the localisation of the gene-trapped KRAB-ZFPs with respect to cellular differentiation and the cell cycle. In differentiated gene-trap ES cells, I found that both KAP-1 and the gene-trapped KRAB-ZFPs relocate to pericentromeric heterochromatin, and that KAP-1 is probably required, but is not sufficient, for this localisation. This finding was in agreement with Cammas *et al.* (2002) who also showed a significant re-localisation of KAP-1 to heterochromatin in ES and F9 cells after 48 hr of RA differentiation. That the distribution of KAP-1 alters with cell differentiation was not unexpected, given its known role as a co-repressor during differentiation and development (Cammas *et al.*, 2000, Weber *et al.*, 2002) but it is interesting to note that KAP-1 is also required as a co-activator for induced differentiation of the myelomonocytic cell line U937 (Rooney and Calame, 2001). It would therefore be interesting to study whether KAP-1 re-locates to heterochromatin in this cell line after differentiation. Heterochromatin association of the gene-trapped KRAB-ZFPs and KAP-1, were also found to correlate with changes

in proliferation of cells but not with any particular stage of the cell cycle, which was also in agreement with Cammas *et al.* (2002).

I furthered the investigation into KRAB-ZFP localisation by cloning KRAB-ZFP 492 (Chapter 3), generating full-length and mutant GFP tagged constructs and raising an antibody to the endogenous protein (Chapter 4). Neither the KRAB nor ZF domains are solely responsible for nuclear targeting, but the GFP-tagged constructs confirmed that the KRAB-domain is required for localisation to heterochromatin. Full-length GFP-tagged 492 was found at other discrete foci in the nucleoplasm suggesting the formation of such foci require the ZF domain. By antibody staining I confirmed that endogenous 492 protein (Zfp647) is present in discrete euchromatic foci but no concentration of protein was observed at pericentromeric heterochromatin. Moreover, endogenous KRAB-ZFP 492 was found in nucleoli of undifferentiated cells, and re-localised to euchromatic foci in a sub-population of differentiated ES cells. These foci often co-localised with KAP-1 and HP1 proteins (Chapter 5). A similar sub-nuclear localisation pattern was seen for another endogenous KRAB-ZFP, NT-2, in differentiating ATDC5 cells (Chapter 5).

The production of antibodies to endogenous KRAB-ZFP 492 has enabled a more thorough study of KRAB-ZFP distribution and the fact that this antibody can also recognise its human homologue will prove useful for future functional studies. These localisation studies suggest that epitope tagging of KRAB-ZFPs may lead to erroneous results. For example, none of the gene-trapped or GFP-tagged 492 proteins were enriched in nucleoli. In both cases the tag may interrupt associations with other nucleolar proteins and/or RNA/DNA causing them to mis-localise. A potential role for nucleolar localisation of KRAB-ZFP 492 and NT-2 in the nucleolus may be for storage, or because it has RNA or DNA targets in this location. Proteins are often stored in the nucleolus and released for later use. In yeast, cell cycle-regulation is linked to nucleolar sequestration. For example, *CDC14* belongs to a group of genes that have a critical role in exit from mitosis. During G1, Cdc14 is inactive and localises to the nucleolus. Following entry into mitosis and specifically during anaphase, Cdc14 is liberated from its nucleolar binding partner, Net1/Cfi1,

and becomes active (reviewed by Carmo-Fonseca *et al.*, 2002). Section 4.9 gives more examples.

What determines the localisation of KRAB-ZFPs in the cell and particularly their localisation in KAKA foci? The linker region and ZF domains of KRAB-ZFPs may also contribute to the sub-cellular localisation of KRAB-ZFPs. RBaK interacts with pRB via its linker region (Skapek *et al.*, 2000), as does NRIF with $p75^{NTR}$ (Casademunt *et al.*, 1999). In addition to protein interactions mediated by the linker and ZF domain, the ZF domain may also be responsible for DNA interactions. Both of these would be expected to affect KRAB-ZFP distribution and in transfected cells, the ZF domain of KRAB-ZFP 492 is required for the formation of KAKA foci. Therefore, foci of NT-2 and 492 may be a direct representation of protein or DNA interactions mediated via the ZF domain. There is evidence to suggest that either of these models is plausible. KAKA foci may represent 'silencing factories' determined by the clustering of repressed genes by a common KRAB-ZFP, similar to the 'transcription factories' described by Osborne *et al.*, 2004. To prove this I will need to show that target genes, which are repressed, are localised at these factories. Otherwise, KAKA foci might represent a sub-nuclear compartment where KRAB-ZFPs, along with other proteins they interact with, are either stored or used in a different facet of transcriptional regulation. However, in the case of ZBrK1, the same ZFs are required for interaction with BRCA1, homo-oligomerisation and DNA binding (Tan *et al.*, 2004a; Tan *et al.*, 2004b). Therefore competition for this binding domain might therefore play a role in its functional localisation. Competition for binding sites has also been postulated for NRIF where it competes with TRAF6 for binding to $p75^{NTR}$ (Casademunt *et al.*, 1999) and this directly affects its sub-cellular distribution (Gentry *et al.*, 2004). If such competition were to occur between proteins that bind to the KRAB domain, for example KAP-1 and pVHL (Li *et al.*, 2003) or KAP-1 and TRAF6 (Gentry *et al.*, 2004) then this too may affect both KRAB-ZFP localisation and also the extent of target gene repression.

7.4 The mechanism of KRAB-KAP-1 mediated repression

The euchromatic 492 foci also contain KAP-1 and HP1 proteins and are likely to mark heterochromatic-like structures involved in KRAB-KAP-1-HP1 mediated gene repression (Chapter 5). This is in broad agreement with Ayyanathen *et al.*, 2003. By using mutant ES cells, I discovered the formation of KAKA foci is independent of the HMT Suv39h1/2. However, the localisation of KAP-1 at pericentromeric heterochromatin is Suv39h1/2 dependent but independent of DNA methylation by Dnmt3a/b (Chapter 5). This suggests the permanent inactivation of genes by KAP-1 at heterochromatin does not necessarily involve their DNA methylation and calls into question the role of KAP-1 at heterochromatin. Therefore, together with current evidence, I predict that a KAP-1 recruits the HMT, SETDB1, to methylate histone H3-K9 residues as a prerequisite to HP1 binding and the subsequent formation of discrete KAKA foci (Schultz *et al.*, 2002; Ayyanathen *et al.*, 2003). In the future it would be interesting to investigate the localisation of SETDB1 (ESET) with respect to these foci. There is no mutant SETDB1 ES cell line reported to date, but it would be interesting to deplete this protein by RNAi, and then examine the sub-nuclear localisation of KRAB-ZFPs and KAP-1. In order for SETDB1 to methylate K9 residues, deacetylation of K9 must first occur. KAP-1 interacts with Mi-2α of the NuRd HDAC complex (Schultz *et al.*, 2001) and has also been isolated in complex with N-CoR-1 and HDAC3 (Underhill *et al.*, 2000), although KAP-1 is only in a complex with these proteins some of the time. Nevertheless, the mechanism of KAP-1-KRAB-ZFP repression may involve histone deacetylation since TSA can interfere this repression (Matsuda *et al.*, 2001). Therefore it would be interesting to see if KAKA foci are dissociated by TSA treatment.

My results support the idea that the consistent coupling of chromatin remodelling enzymes with KRAB-ZFPs results in localised chromatin changes in the vicinity of target genes, rather than translocation of KRAB-ZFP target genes to pericentromeric heterochromatin for KRAB-KAP-1 mediated repression (Fig. 5.1B). The fact that Ayyanathen *et al.*, (2003) reported translocation of a luciferase transgene to

heterochromatin may be because it is a transgene and not an endogenous locus. It is also important not to forget the role of active repression by the KRAB-KAP-1 complex. KRAB-KAP-1 repression is efficient in the short term (Pengue *et al.*, 1996; Agata *et al.*, 1999) and repression is specific to RNA polymerase II and III reporters, suggesting that KRAB-KAP-1 may interfere with some components of the transcription machinery (Moosmann *et al.*, 1997) (section 1.1.5.2).

In this thesis I focused on KRAB-ZFP repression mediated by interaction with the co-repressor KAP-1, but it is important to remember that KRAB-ZFPs can also interact independently with other co-repressors, some of which contain protein motifs similar to KAP-1 such as RING or PHD motifs. BRCA1 contains a RING motif at the N-terminus but interaction occurs with ZBrK1 via its C-terminal domain. ZBrK1 thus has two repression domains, an N-terminal KRAB domain and a BRCA1 interaction domain which functions in a HDAC and promoter-specific manner (Tan *et al.*, 2004a). NSD1 is a HMT that specifically methylates histone H3 at K36 and histone H4 at K20 (Rayasam *et al.*, 2003) via its SET domain, but it also contains five PHD domains. Interaction with KRAB-ZFP Nizp1 is via a cysteine-rich motif and when tethered to an RNA polymerase II promoter repression is NSD1 dependent (Nielsen *et al.*, 2004). These interactions suggest that KRAB-ZFPs are involved in the sequence specific repression of genes by a number of different mechanisms that depend on interactions with various co-repressors, and that they may even function as a bridging protein, linking one repression pathway to the next.

7.5 KRAB-ZFP transgenic mice

When the 113 and 492 KRAB-ZFP genes are disrupted, the phenotype in mice appears to be homozygous lethal, based on the number of offspring resulting from het x het matings. Both expression patterns in mice point towards a role in embryonic development and such expression patterns are common within the KRAB-ZFP family (section 6.5). There is speculation as to why there are so many KRAB-ZFPs in mammals and whether they have redundant functions during development. I generated two transgenic mice whose mutated gene was situated in different genomic environments. Although I do not know the identity of the gene

encoding KRAB-ZFP 113, I anticipated that the phenotype of this knock-out might not be as severe as the 492 knock-out. KRAB-ZFP 492 has homologues in human and rat. However, although there are many KRAB-ZFP genes in this cluster, 113 appears to be essential for development. In the future, this mouse may be useful in addressing the question of redundancy at this locus.

Since mammals have substantially different repertoires of KRAB-ZFPs it has been suggested that as these genes evolve, they acquire novel functions over evolutionary time and this contributes to species-specific development (Shannon *et al.*, 2003). The presence of TIF1 orthologues in lower animals reflects the emergence of KRAB-ZFPs. TIF1α and/or -γ orthologues have been found in *Drosophila*, *Xenopus*, and zebrafish (Beckstead *et al.*, 2001; Ransom *et al.*, 2004), but so far no orthologues of KAP-1 have been found in zebrafish or *fugu*. Homologues of the KRAB-domain have neither been found in zebrafish nor *fugu*, although they have been found in many species of frog but not in *Drosophila* (Shannon *et al.*, 2003). I found that gene-trapped KRAB-ZFPs did not localise to heterochromatin unless KAP-1 was present (Chapter 4), suggesting that KAP-1 was the initial member of this partnership. It has been proposed that KAP-1 brings together different KRAB-ZFPs under a similar mechanism of KAP-1-mediated control, and this is consistent with the idea that the evolution of body-plans, especially among mammals, is mostly due to remodelling of regulatory circuits that control gene expression patterns (Carroll, 1995). Therefore, an explanation for the large number of essential, non-redundant KRAB-ZFPs required in mammals is to allow for the complexity of their body-plans, and the KRAB domain permits for their concerted control by KAP-1 (Shannon *et al.*, 2003).

7.6 Future directions

Comparative analysis of mouse and human draft genomes reveals that ~99% of mouse genes have an orthologue in man, underscoring the importance of studying gene function in mouse (Krebs *et al.*, 2003 and reference therein). The remaining 1% of genes includes differentially expanded members of multi-gene families, which function in reproduction, xenobiotic metabolism, and immunity. This complex group

of genes may encode functions either related to speciation, or encode the regulators of such genes. Their individual characterisation and functional importance are usually more difficult to assess experimentally. The majority of KRAB-ZFPs fall into this latter class of genes and despite the advances made in this area of research during the course of my PhD; there remains an abundance of questions to be answered about this protein family.

Perhaps the most intriguing question of all is that of speciation. What makes us human after all? This question can only be answered by the study of protein families such as the KRAB-ZFP family. In this respect, further analysis of the genes identified in this thesis will provide a greater selection of KRAB-ZFP genes with various evolutionary backgrounds that can be studied from a functional perspective by taking advantage of the gene-trap system. The difficulty of working with this group of proteins is twofold however, and should be considered for future experimentation. Firstly, their sequence similarity can be a factor of inconvenience (NT-2, section 5.3) and endogenous proteins should always be tested if possible.

The little information we have on transcriptional repression by KRAB-ZFPs stems from a gap in knowledge derived from the limited availability of KRAB-ZFP target sequences. This will undoubtedly be the next hurdle to cross with respect to the future direction of this research area. The mouse model generated in this thesis will be of considerable help in identifying the targets of KRAB-ZFP 492 as well as addressing the more functional questions such as the impact of an abnormal KAP-1-KRAB-ZFP 492 interaction during embryonic differentiation. However, the presumed role of KRAB-ZFPs in sequence-specific DNA binding is unlikely to be their only mode of gene repression. Protein-protein interactions are likely to play an indirect but significant role. To my knowledge, KRAB-ZFPs have been reported to interact with at least 10 other proteins, excluding KAP-1 and viral proteins (Liao *et al.*, 2005). Intriguingly, six of these are either tumour suppressors or oncogenes and the interacting KRAB-ZFPs have been isolated and characterised because of their association with them (Casademunt *et al.*, 1999; Zheng *et al.*, 2000; Skapek *et al.*, 2000; Li *et al.*, 2003; Hennemann *et al.*, 2003; Nielsen *et al.*, 2004). On this

premise, it would be interesting to identify protein interactors of KRAB-ZFP 492 by yeast-two hybrid assays for example and such experiments will provide clues as to the pathways individual KRAB-ZFPs are involved in.

Many KRAB-ZFPs are known to have direct roles in human disease (Wagner *et al.*, 2000; Alders *et al.*, 2000; Garcia *et al.*, 2004) and many more are implicated in disease-associated pathways (Skapek *et al.*, 2000; Li *et al.*, 2003; Nielsen *et al.*, 2004). Surprisingly and despite their overwhelming numbers, this thesis (Chapter 6) and previous evidence points to the sum of four KRAB-ZFPs being critically important for embryonic development by virtue of mouse models. No functional redundancy has been shown as yet, despite the large number of similar KRAB-ZFPs in the genome. Therefore, we should continue to study this protein family in such a way, firstly to understand in full the functional implications of a given KRAB-ZFP, but secondly for the potential discovery of new human disease causing genes.

Finally, I predict that KRAB-ZFPs will be major players in the linking and switching of different biological networks throughout differentiation and development. Alternative splicing and their ability to bind both protein and/or DNA gives them this functional capacity. Interaction with chromatin-associated proteins also indicates a link between various repression pathways and the temporary and permanent inactivation of genes. An important future goal with respect to this work will be to analyse the dynamics of a KRAB-ZFP target gene in the context of the model proposed in this thesis as not all KRAB-ZFPs are likely to behave the same way-or are they? Many more need to be studied in order to assess whether the formation of endogenous foci are a common attribute. Surely KRAB-ZFPs are an important and essential area of future research, as the ultimate goal of molecular biology is to not only map the intricate gene-regulatory processes of the human body but to assemble them into one pathway and KRAB-ZFPs could be the missing link.

References

Aagaard,L., Laible,G., Selenko,P., Schmid,M., Dorn,R., Schotta,G., Kuhfittig,S., Wolf,A., Lebersorger,A., Singh,P.B., Reuter,G., Jenuwein,T. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. *18*, 1923-1938.

Agalioti T, Chen G, Thanos D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. Cell. *111*, 381-92.

Aasland,R., Gibson,T.J., Stewart,A.F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci. *20*, 56-9.

Åbrink M, Ortiz JA, Mark C, Sanchez C, Looman C, Hellman L, Chambon P, Losson R (2001). Conserved interaction between distinct Kruppel-associated box domains and the transcriptional intermediary factor 1 beta. Proc.Natl.Acad.Sci.U.S.A. *98*, 1422-1426.

Agata,Y., Matsuda,E., Shimizu,A. (1999). Two novel Kruppel-associated box-containing zinc-finger proteins, KRAZ1 and KRAZ2, repress transcription through functional interaction with the corepressor KAP-1 (TIF1beta/KRIP-1). J.Biol.Chem. *274*, 16412-16422.

Ahringer J (2000). NuRD and SIN3 histone deacetylase complexes in development. Trends.Genet. *16*, 351-356.

Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K., Watson,J.D. Molecular Biology of the Cell. [3rd Edition]. 1994. Garland Publishing.

Alders,M., Ryan,A., Hodges,M., Bliek,J., Feinberg,A.P., Privitera,O., Westerveld,A., Little,P.F., Mannens,M. (2000). Disruption of a novel imprinted zinc-finger gene, ZNF215, in Beckwith-Wiedemann syndrome. Am.J.Hum.Genet. *66*, 1473-84.

Alonso,M.B., Zoidl,G., Taveggia,C., Bosse,F., Zoidl,C., Rahman,M., Parmantier,E., Dean,C.H., Harris,B.S., Wrabetz,L., Muller,H.W., Jessen,K.R., Mirsky,R. (2004). Identification and characterization of ZFP-57, a novel zinc finger transcription factor in the mammalian peripheral nervous system. J.Biol.Chem. *279*, 25653-25664.

Andersen,J.S., Lyon,C.E., Fox,A.H., Leung,A.K., Lam,Y.W., Steen,H., Mann,M., Lamond,A.I. (2002). Directed proteomic analysis of the human nucleolus. Curr Biol. *12*, 1-11.

Andrulis ED, Zappulla DC, Ansari A, Perrod S, Laiosa CV, Gartenberg MR, Sternglanz R.(20020. Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning.
Mol Cell Biol. *22*, 8292-301.

Armstrong SJ, Hulten MA, Keohane AM, Turner BM.(1997). Different strategies of X-inactivation in germinal and somatic cells: histone H4 underacetylation does not mark the inactive X chromosome in the mouse male germline.Exp Cell Res. *230*, 399-402.

Ayyanathan,K., Lechner,M.S., Bell,P., Maul,G.G., Schultz,D.C., Yamada,Y., Tanaka,K., Torigoe,K., Rauscher,F.J., III (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. *17*, 1855-1869.

Bachman,K.E., Rountree,M.R., Baylin,S.B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. J.Biol.Chem. *276*, 32282-32287.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. *410*, 120-124.

Barlow,P.N., Luisi,B., Milner,A., Elliott,M., Everett,R. (1994). Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. J.Mol.Biol. *237*, 201-11.

Beckstead,R.B., Ner,S.S., Hales,K.G., Grigliatti,T.A., Baker,B.S., Bellen,H.J. (2004). Bonus, a Drosophila TIF1 homologue, is a chromatin associated protein that acts as a modifier of position effect variegation. Genetics. [Epub ahead of print]

Beckstead,R., Ortiz,J.A., Sanchez,C., Prokopenko,S.N., Chambon,P., Losson,R., Bellen,H.J. (2001). Bonus, a Drosophila homolog of TIF1 proteins, interacts with nuclear receptors and can inhibit betaFTZ-F1-dependent transcription. Mol.Cell. *7*, 753-65.

Beisel,C., Imhof,A., Greene,J., Kremmer,E., Sauer,F. (2002). Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature. *419*, 857-62.

Bell,P., Lieberman,P.M., Maul,G.G. (2000). Lytic but not latent replication of epstein-barr virus is associated with PML and induces sequential release of nuclear domain 10 proteins. J.Virol. *74*, 11800-11810.

Bellefroid EJ, Marine JC, Matera AG, Bourguignon C, Desai T, Healy KC, Bray-Ward P, Martial JA, Ihle JN, Ward DC (1995). Emergence of the ZNF91 Kruppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. Proc.Natl.Acad.Sci.U.S.A. *92*, 10757-10761.

Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991). The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc.Natl.Acad.Sci.U.S.A. *88*, 3608-3612.

Bellefroid EJ, Marine JC, Ried T, Lecocq PJ, Riviere M, Amemiya C, Poncelet DA, Coulie PG, de Jong P, Szpirer C, . (1993). Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. EMBO.J. *12*, 1363-1374.

Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP, Allshire RC. (2001). Requirement of heterochromatin for cohesion at centromeres. Science. *294*, 2539-42.

Bernardi,R., Scaglioni,P.P., Bergmann,S., Horn,H.F., Vousden,K.H., Pandolfi,P.P. (2004). PML regulates p53 stability by sequestering Mdm2 to the nucleolus. Nat.Cell Biol. *6*, 665-672.

Bestor,T., Laudano,A., Mattaliano,R., Ingram,V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J.Mol.Biol. *203*, 971-983.

Bickmore,W.A., Sutherland,H.G. (2002). Addressing protein localization within the nucleus. EMBO J. *21*, 1248-1254.

Billon,N., Jolicoeur,C., Ying,Q.L., Smith,A., Raff,M. (2002). Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells. J.Cell Sci. *115*, 3657-3665.

Bird AP, Wolffe AP.(1999). Methylation-induced repression--belts, braces, and chromatin. Cell. *99*, 451-4.

Blindauer,C.A., Harrison,M.D., Parkinson,J.A., Robinson,A.K., Cavet,J.S., Robinson,N.J., Sadler,P.J. (2001). A metallothionein containing a zinc finger within a four-metal cluster protects a bacterium from zinc toxicity. Proc.Natl.Acad.Sci.U.S.A. *98*, 9593-8.

Boeke,J., Ammerpohl,O., Kegel,S., Moehren,U., Renkawitz,R. (2000). The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J.Biol.Chem. 275, 34963-7.

Boggs BA, Cheung P, Heard E, Spector DL, Chinault AC, Allis CD.(2002). Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. Nat.Genet. *30*, 73-6.

Borden,K.L., Boddy,M.N., Lally,J., O'Reilly,N.J., Martin,S., Howe,K., Solomon,E., Freemont,P.S. (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. EMBO J. *14*, 1532-41.

Borden,K.L., Martin,S.R., O'Reilly,N.J., Lally,J.M., Reddy,B.A., Etkin,L.D., Freemont,P.S. (1993). Characterisation of a novel cysteine/histidine-rich metal binding domain from Xenopus nuclear factor XNF7.FEBS Lett. *335*, 255-60.

Boyle,S., Gilchrist,S., Bridger,J.M., Mahy,N.L., Ellis,J.A., Bickmore,W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum.Mol.Genet. *10*, 211-219.

Bridger JM, Kill IR, Lichter P. (1998). Association of pKi-67 with satellite DNA of the human genome in early G1 cells. Chromosome Res. *6*, 13-24.

Brown,K.E., Guest,S.S., Smale,S.T., Hahm,K., Merkenschlager,M., Fisher,A.G. (1997). Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. Cell *91*, 845-854.

Brown,K.E., Baxter,J., Graf,D., Merkenschlager,M., Fisher,A.G. (1999). Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. Mol.Cell *3*, 207-217.

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD.(1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation.Cell. *84*, 843-51.

Bruno S, Ghiotto F, Fais F, Fagioli M, Luzi L, Pelicci PG, Grossi CE, Ciccone E. (2003). The PML gene is not involved in the regulation of MHC class I expression in human cell lines. Blood. *101*, 3514-9.

Bulger M, Schubeler D, Bender MA, Hamilton J, Farrell CM, Hardison RC, Groudine M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. Mol.Cell.Biol. *23*, 5234-44.

Bultman,S., Gebuhr,T., Yee,D., La Mantia,C., Nicholson,J., Gilliam,A., Randazzo,F., Metzger,D., Chambon,P., Crabtree,G., Magnuson,T. (2000). A Brg1 null mutation in the

mouse reveals functional differences among mammalian SWI/SNF complexes. Mol.Cell. *6*, 1287-95.

Bulyk,M.L., Huang,X., Choo,Y., Church,G.M. (2001). Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc.Natl.Acad.Sci.U.S.A *98*, 7158-7163.

Cammas F, Mark M, Dolle P, Dierich A, Chambon P, Losson R (2000). Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. Development. *127*, 2955-2963.

Cammas F, Oulad-Abdelghani M, Vonesch JL, Huss-Garcia Y, Chambon P, Losson R (2002). Cell differentiation induces TIF1beta association with centromeric heterochromatin via an HP1 interaction. J.Cell.Sci. *115* , 3439-3448.
Cammas,F., Herzog,M., Lerouge,T., Chambon,P., Losson,R. (2004). Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. Genes Dev. *18*, 2147-2160.

Cao,R., Wang,L., Wang,H., Xia,L., Erdjument-Bromage,H., Tempst,P., Jones,R.S., Zhang Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-43.
Caricasole,A., Duarte,A., Larsson,S.H., Hastie,N.D., Little,M., Holmes,G., Todorov,I., Ward,A. (1996). RNA binding by the Wilms tumor suppressor zinc finger proteins. Proc.Natl.Acad.Sci.U.S.A. *93*, 7562-6.

Carmo-Fonseca M (2002). The contribution of nuclear compartmentalization to gene regulation. Cell. *108*, 513-521.

Carmo-Fonseca M, Mendes-Soares L, Campos I (2000). To be or not to be in the nucleolus. Nat.Cell.Biol. *2*, E107-E112.

Carroll,S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. Nature. *376*, 479-85.

Carruthers,L.M., Bednar,J., Woodcock,C.L., Hansen,J.C. (1998). Linker histones stabilize the intrinsic salt-dependent folding of nucleosomal arrays: mechanistic ramifications for higher-order chromatin folding. Biochemistry. *37*, 14776-87.

Casademunt,E., Carter,B.D., Benzel,I., Frade,J.M., Dechant,G., Barde,Y.A. (1999). The zinc finger protein NRIF interacts with the neurotrophin receptor p75(NTR) and participates in programmed cell death. EMBO J. *18*, 6050-6061.

Cavalli G, Paro R. (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. Science. *286*, 955-8.

Chadwick,B.P., Willard,H.F. (2001). Histone H2A variants and the inactive X chromosome: identification of a second macroH2A variant. Hum.Mol.Genet. *10*, 1101-1113.

Chakrabarti SK, Francis J, Ziesmann SM, Garmey JC, Mirmira RG.(2003). Covalent histone modifications underlie the developmental regulation of insulin gene transcription in pancreatic beta cells. J.Biol.Chem. *278*, 23617-23.

Chambeyron,S., Bickmore,W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. *18*, 1119-1130.

Chang,C.J., Chen,Y.L., Lee,S.C. (1998). Coactivator TIF1beta interacts with transcription factor C/EBPbeta and glucocorticoid receptor to induce alpha1-acid glycoprotein gene expression. Mol.Cell Biol. *18*, 5880-5887.

Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR.(1999). Regulation of transcription by a protein methyltransferase. Science. *284*, 2174-7.

Chen,T., Ueda,Y., Dodge,J.E., Wang,Z., Li,E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol.Cell Biol. *23*, 5594-5605.

Chen,T., Tsujimoto,N., Li,E. (2004). The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. Mol.Cell Biol. *24*, 9048-9058.

Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell. *5*, 905-15.

Chubb,J.R., Bickmore,W.A. (2003). Considering nuclear compartmentalization in the light of nuclear dynamics. Cell *112*, 403-406.

Clemson,C.M., Chow,J.C., Brown,C.J., Lawrence,J.,B. (1998). Stabilization and localization of Xist RNA are controlled by separate mechanisms and are not sufficient for X inactivation. J Cell Biol. *142*, 13-23.

Cohen-Armon M, Visochek L, Katzoff A, Levitan D, Susswein AJ, Klein R, Valbrun M, Schwartz JH. (2004). Long-term memory requires polyADP-ribosylation. Science. *304*, 1820-2.

Collins,T., Stone,J.R., Williams,A.J. (2001). All in the family: the BTB/POZ, KRAB, and SCAN domains. Mol.Cell Biol. *21*, 3609-3615.

Cook,P.R. (2003). Nongenic transcription, gene regulation and action at a distance. J.Cell Sci. *116*, 4483-4491.

Côté F, Boisvert FM, Grondin B, Bazinet M, Goodyer CG, Bazett-Jones DP, Aubry M (2001). Alternative promoter usage and splicing of ZNF74 multifinger gene produce protein isoforms with a different repressor activity and nuclear partitioning. DNA.Cell.Biol. *20*, 159-173.

Cirillo,L.A., McPherson,C.E., Bossard,P., Stevens,K., Cherian,S., Shim,E.Y., Clark,K.L., Burley,S.K., Zaret,K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244-54.

Craig,J.M., Bickmore,W.A. (1994). The distribution of CpG islands in mammalian chromosomes. Nat.Genet. *7*, 376-82.

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J.Cell.Biol. *145*, 1119-1131.

Crossley,P.H., Little,P.F. (1991). A cluster of related zinc finger protein genes is deleted in the mouse embryonic lethal mutation tw18. Proc.Natl.Acad.Sci.U.S.A. *88*, 7923-7.

Csink,A.K., Henikoff,S. (1996). Genetic modification of heterochromatic association and nuclear organization in Drosophila. Nature *381*, 529-531.

Dai,J., Li,Y., Ji,C., Jin,F., Zheng,Z., Wang,X., Sun,X., Xu,X., Gu,S., Xie,Y., Mao,Y. (2003). Characterization of two novel KRAB-domain-containing zinc finger genes, ZNF460 and ZNF461, on human chromosome 19q13.1-->q13.4. Cytogenet.Genome Res. *103*, 74-78.

Dechat T, Korbei B, Vaughan OA, Vlcek S, Hutchison CJ, Foisner R. (2000). Lamina-associated polypeptide 2alpha binds intranuclear A-type lamins. J.Cell.Sci. *113*, 3473-84.

del Mar Lorente,M., Marcos-Gutierrez,C., Perez,C., Schoorlemmer,J., Ramirez,A., Magin,T., Vidal,M. (2000). Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice.Development. *127*, 5093-100.

De Lucchini,S., Rijli,F.M., Ciliberto,G., Barsacchi,G. (1991). A Xenopus multifinger protein, Xfin, is expressed in specialized cell types and is localized in the cytoplasm. Mech.Dev. *36*, 31-40.

de Napoles,M., Mermoud,J.E., Wakao,R., Tang,Y.A., Endoh,M., Appanah,R., Nesterova,T.B., Silva,J., Otte,A.P., Vidal,M., Koseki,H., Brockdorff,N. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev.Cell. *7*, 663-76.

Dehal,P., Predki,P., Olsen,A.S., Kobayashi,A., Folta,P., Lucas,S., Land,M., Terry,A., Ecale Zhou,C.L., Rash,S., Zhang,Q., Gordon,L., Kim,J., Elkin,C., Pollard,M.J., Richardson,P., Rokhsar,D., Uberbacher,E., Hawkins,T., Branscomb,E., Stubbs,L. (2001). Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. Science *293*, 104-111.

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM.(1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature.*399*, 491-6.

Dillon N, Festenstein R. (2002). Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. Trends Genet. *18*, 252-8.

Dodge,J.E., Kang,Y.K., Beppu,H., Lei,H., Li,E. (2004). Histone H3-K9 Methyltransferase ESET Is Essential for Early Development. Mol.Cell Biol. *24*, 2478-2486.

Dreyer SD, Zhou L, Machado MA, Horton WA, Zabel B, Winterpacht A, Lee B (1998). Cloning, characterization, and chromosomal assignment of the human ortholog of murine Zfp-37, a candidate gene for Nager syndrome. Mamm.Genome. *9*, 458-462.

Dreyer SD, Zheng Q, Zabel B, Winterpacht A, Lee B (1999) Isolation, characterization, and mapping of a zinc finger gene, ZFP95, containing both a SCAN box and an alternatively spliced KRAB A domain. Genomics. *62*, 119-122.

Dundr M, Misteli T (2001). Functional architecture in the cell nucleus. Biochem.J. *356*, 297-310.

Elbashir SM, Lendeckel W, Tuschl T.(2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188-200.

Ebert,A., Schotta,G., Lein,S., Kubicek,S., Krauss,V., Jenuwein,T., Reuter,G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev. *18*, 2973-2983.

Eichler,E.E., Hoffman,S.M., Adamson,A.A., Gordon,L.A., McCready,P., Lamerdin,J.E., Mohrenweiser,H.W. (1998). Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. Genome Res. *8*, 791-808.

Eissenberg,J.C., Morris,G.D., Reuter,G., Hartnett,T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics *131*, 345-352.

Espada,J., Ballestar,E., Fraga,M.F., Villar-Garea,A., Juarranz,A., Stockert,J.C., Robertson,K.D., Fuks,F., Esteller,M. (2004). Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J.Biol.Chem. *279*, 37175-37184.

Fairall,L., Schwabe,J.W., Chapman,L., Finch,J,T., Rhodes,D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature. *366*, 483-7.

Festenstein,R., Sharghi-Namini,S., Fox,M., Roderick,K., Tolaini,M., Norton,T., Saveliev,A., Kioussis,D., Singh,P. (1999). Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner. Nat.Genet. *23*, 457-461.

Festenstein,R., Pagakis,S.N., Hiragami,K., Lyon,D., Verreault,A., Sekkali,B., Kioussis,D. (2003). Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells. Science *299*, 719-721.

Firestein,R., Cui,X., Huie,P., Cleary,M.L. (2000). Set domain-dependent regulation of transcriptional silencing and growth control by SUV39H1, a mammalian ortholog of Drosophila Su(var)3-9. Mol.Cell Biol. *20* , 4900-4909.

Fischle,W., Wang,Y., Allis,C.D. (2003a). Binary switches and modification cassettes in histone biology and beyond. Nature *425*, 475-479.

Fischle,W., Wang,Y., Allis,C.D. (2003b). Histone and chromatin cross-talk. Curr. Opin. Cell Biol. *15*, 172-183.

Frade,J.M., Barde,Y.A. (1999). Genetic evidence for cell death mediated by nerve growth factor and the neurotrophin receptor p75 in the developing mouse retina and spinal cord. Development *126*, 683-690.

Fraser,R.A., Heard,D.J., Adam,S., Lavigne,A.C., Le Douarin,B., Tora,L., Losson,R., Rochette-Egly,C., Chambon,P. (1998). The putative cofactor TIF1alpha is a protein kinase that is hyperphosphorylated upon interaction with liganded nuclear receptors. J.Biol.Chem. *273*, 16199-16204.

Frey MR, Bailey AD, Weiner AM, Matera AG. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. Curr.Biol. *9*, 126-35.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ 3rd (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes.Dev. *10*, 2067-2078.

Fujita,N., Watanabe,S., Ichimura,T., Tsuruzoe,S., Shinkai,Y., Tachibana,M., Chiba,T., Nakao,M. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J.Biol.Chem. *278*, 24132-24138.

Fuks,F., Hurd,P.J., Deplus,R., Kouzarides,T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. *31*, 2305-2312.

Galy V, Olivo-Marin JC, Scherthan H, Doye V, Rascalou N, Nehrbass U. (2000). Nuclear pore complexes in the organization of silent telomeric chromatin. Nature. *403*, 108-12.

Garcia,V., Dominguez,G., Garcia,J.M., Silva,J., Pena,C., Silva,J.M., Carcereny,E., Menendez,J., Espana,P., Bonilla,F. (2004). Altered expression of the ZBRK1 gene in human breast carcinomas. J.Pathol. *202*, 224-232.

Gartenberg MR.(2000). The Sir proteins of Saccharomyces cerevisiae: mediators of transcriptional silencing and much more. Curr.Opin.Microbiol. *3*, 132-7.

Gehring,W.J. (1985). The homeo box: a key to the understanding of development? Cell. *40*, 3-5.

Gentry,J.J., Rutkoski,N.J., Burke,T.L., Carter,B.D. (2004). A functional interaction between the p75 neurotrophin receptor interacting factors, TRAF6 and NRIF. J.Biol.Chem. *279*, 16646-16656.

Gilbert,N., Boyle,S., Sutherland,H., de Las,H.J., Allan,J., Jenuwein,T., Bickmore,W.A. (2003). Formation of facultative heterochromatin in the absence of HP1. EMBO J. *22*, 5540-5550.

Gilbert,N., Boyle,S., Fiegler,H., Woodfine,K., Carter,N.P., Bickmore,W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555-566.

Gilchrist,S., Gilbert,N., Perry,P., Bickmore,W.A. (2004). Nuclear organization of centromeric domains is not perturbed by inhibition of histone deacetylases. Chromosome.Res. *12*, 505-516.

Goldmark JP, Fazzio TG, Estep PW, Church GM, Tsukiyama T.(2000). The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. Cell. *103*, 423-33.

Goldmit M, Ji Y, Skok J, Roldan E, Jung S, Cedar H, Bergman Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. Nat Immunol. *6*,198-203.

Gou,D., Wang,J., Gao,L., Sun,Y., Peng,X., Huang,J., Li,W. (2004). Identification and functional analysis of a novel human KRAB/C2H2 zinc finger gene ZNF300. Biochim.Biophys.Acta *1676*, 203-209.

Grant PA, Duggan L, Cote J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. 11, 1640-50.

Gravel S, Larrivee M, Labrecque P, Wellinger RJ.(1998). Yeast Ku as a regulator of chromosomal DNA end structure. Science. *280*, 741-4.

Grigoryev,S.A., Nikitina,T., Pehrson,J.R., Singh,P.B., Woodcock,C.L. (2004). Dynamic relocation of epigenetic chromatin markers reveals an active role of constitutive heterochromatin in the transition from proliferation to quiescence. J.Cell Sci.

Grondin B, Bazinet M, Aubry M (1996). The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. J.Biol.Chem. *271*, 15458-15467.

Grondin,B., Cote,F., Bazinet,M., Vincent,M., Aubry,M. (1997). Direct interaction of the KRAB/Cys2-His2 zinc finger protein ZNF74 with a hyperphosphorylated form of the RNA polymerase II largest subunit. J.Biol.Chem. *272*, 27877-27885.

Govindan,M.V. (1990). Specific region in hormone binding domain is essential for hormone binding and trans-activation by human androgen receptor. Mol.Endocrinol. *4*, 417-27.

Guenatri,M., Bailly,D., Maison,C., Almouzni,G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J.Cell Biol.

Hall,I.M., Shankaranarayana,G.D., Noma,K., Ayoub,N., Cohen,A., Grewal,S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hamilton AJ, Baulcombe DC. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. *286*, 950-2.

Heard,E., Rougeulle,C., Arnaud,D., Avner,P., Allis,C.D., Spector,D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell *107*, 727-738.

Heinzel T, Lavinsky RM, Mullen TM, Soderstrom M, Laherty CD, Torchia J, Yang WM, Brard G, Ngo SD, Davie JR, Seto E, Eisenman RN, Rose DW, Glass CK, Rosenfeld MG. (1997). A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression.
Nature. *387*, 43-8.

Hendrich B, Bird A (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol.Cell.Biol. *18*, 6538-6547.

Hendrich,B., Hardeland,U., Ng,H.H., Jiricny,J., Bird,A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature. *401*, 301-4.

Hennemann,H., Vassen,L., Geisen,C., Eilers,M., Moroy,T. (2003). Identification of a novel Kruppel-associated box domain protein, Krim-1, that interacts with c-Myc and inhibits its oncogenic activity. J.Biol.Chem. *278*, 28799-28811.

Hirschhorn JN, Brown SA, Clark CD, Winston F.(1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. *6*, 2288-98.

Honer C, Chen P, Toth MJ, Schumacher C (2001). Identification of SCAN dimerization domains in four gene families. Biochim.Biophys.Acta. *1517*, 441-448.

Horsley,D., Hutchings,A., Butcher,G.W., Singh,P.B. (1996). M32, a murine homologue of Drosophila heterochromatin protein 1 (HP1), localises to euchromatin within interphase nuclei and is largely excluded from constitutive heterochromatin. Cytogenet.Cell Genet. *73*, 308-311.

Huang,N., vom,B.E., Garnier,J.M., Lerouge,T., Vonesch,J.L., Lutz,Y., Chambon,P., Losson,R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J. *17*, 3398-3412.

Huang,Z., Philippin,B., O'Leary,E., Bonventre,J.V., Kriz,W., Witzgall,R. (1999). Expression of the transcriptional repressor protein Kid-1 leads to the disintegration of the nucleolus. J.Biol.Chem. *274*, 7640-7648.

Iborra,F.J., Pombo,A., Jackson,D.A., Cook,P.R. (1996). Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. J.Cell Sci. *109 (Pt 6)*, 1427-1436.

Imbalzano,A.N., Kwon,H., Green,M.R., Kingston,R.E. (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. Nature. *370*, 481-5.

Jackson DA, Hassan AB, Errington RJ, Cook PR. (1993). Visualization of focal sites of transcription within human nuclei. EMBO J. *12*, 1059-65.

Jackson,D.A., Iborra,F.J., Manders,E.M., Cook,P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Mol.Biol.Cell *9*, 1523-1536.

Jackson,J.P., Lindroth,A.M., Cao,X., Jacobsen,S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. *416*, 556-60.

Jackson,M., Krassowska,A., Gilbert,N., Chevassut,T., Forrester,L., Ansell,J., Ramsahoye,B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol.Cell Biol. *24*, 8862-8871.

Jacobs,S.A., Taverna,S.D., Zhang,Y., Briggs,S.D., Li,J., Eissenberg,J.C., Allis,C.D., Khorasanizadeh,S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. EMBO J. *20*, 5232-5241.

Jacobson RH, Ladurner AG, King DS, Tjian R.(2000). Structure and function of a human TAFII250 double bromodomain module.Science. *288*, 1422-5.

Jheon,A., Chen,J., Teo,W., Ganss,B., Sodek,J., Cheifetz,S. (2002). Temporal and spatial expression of a novel zinc finger transcription factor, AJ18, in developing murine skeletal tissues. J.Histochem.Cytochem. *50*, 973-982.

Jheon,A.H., Ganss,B., Cheifetz,S., Sodek,J. (2001). Characterization of a novel KRAB/C2H2 zinc finger transcription factor involved in bone development. J.Biol.Chem. *276*, 18282-18289.

Jheon,A.H., Suzuki,N., Nishiyama,T., Cheifetz,S., Sodek,J., Ganss,B. (2003). Characterization of the 5'-flanking region of the rat AJ18 gene. Gene *310*, 203-213.

Kaufman,M.H. The Atlas of Mouse Development. 1992. Academic Press Limited.

Katoh,O., Oguri,T., Takahashi,T., Takai,S., Fujiwara,Y., Watanabe,H. (1998). ZK1, a novel Kruppel-type zinc finger gene, is induced following exposure to ionizing radiation and enhances apoptotic cell death on hematopoietic cells. Biochem.Biophys.Res.Commun. *249*, 595-600.

Kedar,V., McDonough,H., Arya,R., Li,H.H., Rockman,H.A., Patterson,C. (2004). Muscle-specific RING finger 1 is a bona fide ubiquitin ligase that degrades cardiac troponin I. Proc.Natl.Acad.Sci.U.S.A. *101*, 18135-40.

Khetchoumian,K., Teletin,M., Mark,M., Lerouge,T., Cervino,M., Oulad-Abdelghani,M., Chambon,P., Losson,R. (2004). TIF1delta, a novel HP1-interacting member of the transcriptional intermediary factor 1 (TIF1) family expressed by elongating spermatids. J.Biol.Chem. *279*, 48329-41.

Kim J, Bergmann A, Stubbs L (2000). Exon sharing of a novel human zinc-finger gene, ZIM2, and paternally expressed gene 3 (PEG3). Genomics. *64*, 114-118.

Kim,J., Bergmann,A., Wehri,E., Lu,X., Stubbs,L. (2001). Imprinting and evolution of two Kruppel-type zinc-finger genes, ZIM3 and ZNF264, located in the PEG3/USP29 imprinted domain. Genomics *77*, 91-98.

Klug,A., Schwabe,J.W. (1995). Protein motifs 5. Zinc fingers. FASEB J. *9*, 597-604.

Klugbauer,S., Rabes,H.M. (1999). The transcription coactivator HTIF1 and a related protein are fused to the RET receptor tyrosine kinase in childhood papillary thyroid carcinomas. Oncogene. *18*, 4388-93.

Knoepfler & Eisenman (1999). Sin meets NuRD and other tails of repression. Cell. 99, 447-50.

Koken,M.H., Reid,A., Quignon,F., Chelbi-Alix,M.K., Davies,J.M., Kabarowski,J.H., Zhu,J., Dong,S., Chen,S., Chen,Z., Tan,C.C., Licht,J., Waxman,S., de The,H., Zelent,A. (1997). Leukemia-associated retinoic acid receptor alpha fusion partners, PML and PLZF, heterodimerize and colocalize to nuclear bodies. Proc.Natl.Acad.Sci.U.S.A *94*, 10255-10260.

Kosak,S.T., Skok,J.A., Medina,K.L., Riblet,R., Le Beau,M.M., Fisher,A.G., Singh,H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science. *296*, 158-62.

Kozak,M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.

Kraus,W.L., Wong,J. (2002). Nuclear receptor-dependent transcription with chromatin. Is it all about enzymes? Eur.J.Biochem. *269*, 2275-2283.

Krebs,C.J., Larkins,L.K., Price,R., Tullis,K.M., Miller,R.D., Robins,D.M. (2003). Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. Genes Dev. *17*, 2664-2674.

Krebs JE, Kuo MH, Allis CD, Peterson CL.(1999). Cell cycle-regulated histone acetylation required for expression of the yeast HO gene.Genes Dev. *13*, 1412-21.

Krumlauf,R. (1994). Hox genes in vertebrate development. Cell. 78, 191-201.

Kuo MH, Allis CD.(1998). Roles of histone acetyltransferases and deacetylases in gene regulation.Bioessays. *20*, 615-26.

Lachner,M., O'Carroll,D., Rea,S., Mechtler,K., Jenuwein,T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lalonde,J.P., Lim,R., Ingley,E., Tilbrook,P.A., Thompson,M.J., McCulloch,R., Beaumont,J.G., Wicking,C., Eyre,H.J., Sutherland,G.R., Howe,K., Solomon,E., Williams,J.H., Klinken,S.P. (2004). HLS5, a novel RBCC (ring finger, B box, coiled-coil) family member isolated from a hemopoietic lineage switch, is a candidate tumor suppressor. J.Biol.Chem. *279*, 8181-9.

Le Douarin B, You J, Nielsen AL, Chambon P, Losson R (1996a). TIF1alpha: a possible link between KRAB zinc finger proteins and nuclear receptors. J.Steroid.Biochem.Mol.Biol. *65*, 43-50.

Le Douarin B, Nielsen AL, Garnier JM, Ichinose H, Jeanmougin F, Losson R, Chambon P (1996b). A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. EMBO.J. *15*, 6701-6715.

Le Douarin,B., Zechel,C., Garnier,J.M., Lutz,Y., Tora,L., Pierrat,P., Heery,D., Gronemeyer,H., Chambon,P., Losson,R. (1995). The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18. EMBO J. *14*, 2020-2033.

Lechner,M.S., Begg,G.E., Speicher,D.W., Rauscher,F.J., III (2000). Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. Mol.Cell Biol. *20*, 6449-6465.

Lee,M.S., Gippert,G.P., Soman,K.V., Case,D.A., Wright,P.E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science. *245*, 635-7.

Lee,P.L., Gelbart,T., West,C., Adams,M., Blackstone,R., Beutler,E. (1997). Three genes encoding zinc finger proteins on human chromosome 6p21.3: members of a new subclass of the Kruppel gene family containing the conserved SCAN box domain. Genomics *43*, 191-201.

Lehnertz,B., Ueda,Y., Derijck,A.A., Braunschweig,U., Perez-Burgos,L., Kubicek,S., Chen,T., Li,E., Jenuwein,T., Peters,A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr.Biol. *13*, 1192-1200.

Leung,A.K., Andersen,J.S., Mann,M., Lamond,A.I. (2003). Bioinformatic analysis of the nucleolus. Biochem.J. *376*, 553-569.

Li,Q., Herrler,M., Landsberger,N., Kaludov,N., Ogryzko,V.V., Nakatani,Y. and Wolffe,A.P. (1998). Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp7 promoter *in vivo*. EMBO J. 17, 6300-6315.

Li,Q., Imhof,A., Collingwood,T.N., Urnov,F.D. and Wolffe,A.P. (1999). p300 stimulates transcription instigated by ligand-bound thyroid hormone receptor at a step subsequent to chromatin disruption. EMBO J. 18, 5634-5652.

Li,Z., Wang,D., Na,X., Schoen,S.R., Messing,E.M., Wu,G. (2003). The VHL protein recruits a novel KRAB-A domain protein to repress HIF-1alpha transcriptional activity. EMBO J. *22*, 1857-1867.

Lichter,P., Bray,P., Ried,T., Dawid,I.B., Ward,D.C. (1992). Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes. Genomics *13*, 999-1007.

Lin,Y.S., Green,M.R. (1991). Mechanism of action of an acidic transcriptional activator in vitro. Cell. *64*, 971-81.

Lippman,Z., Martienssen,R. (2004). The role of RNA interference in heterochromatic silencing. Nature *431*, 364-370.

Looman,C., Abrink,M., Mark,C., Hellman,L. (2002). KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol.Biol.Evol. *19*, 2118-2130.

Looman,C., Mark,C., Abrink,M., Hellman,L. (2003). MZF6D, a novel KRAB zinc-finger gene expressed exclusively in meiotic male germ cells. DNA Cell Biol. *22*, 489-496.

Looman,C., Hellman,L., Abrink,M. (2004). A novel Kruppel-Associated Box identified in a panel of mammalian zinc finger proteins. Mamm.Genome *15*, 35-40.

Lorch Y, Zhang M, Kornberg RD.(1999). Histone octamer transfer by a chromatin-remodeling complex.Cell. *96*,389-92.

Lorenz,P., Koczan,D., Thiesen,H.J. (2001). Transcriptional repression mediated by the KRAB domain of the human C2H2 zinc finger protein Kox1/ZNF10 does not require histone deacetylation. Biol.Chem. *382*, 637-644.

Lorick,K.L., Jensen,J.P., Fang,S., Ong,A.M., Hatakeyama,S., Weissman,A.M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc.Natl.Acad.Sci.U.S.A. *96*, 11364-9.

Luciani,J.J., Depetris,D., Missirian,C., Mignon-Ravix,C., Metzler-Guillemain,C., Megarbane,A., Moncla,A., Mattei,M.G. (2005). Subcellular distribution of HP1 proteins is altered in ICF syndrome. Eur.J.Hum.Genet. *13*, 41-51.

Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., Richmond,T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. *389*, 251-60.

Luo,K., Yuan,W., Zhu,C., Li,Y., Wang,Y., Zeng,W., Jiao,W., Liu,M., Wu,X. (2002). Expression of a novel Krupple-like zinc-finger gene, ZNF382, in human heart. Biochem.Biophys.Res.Commun. *299*, 606-612.

Mahy NL, Perry PE, Bickmore WA. (2002a). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. J.Cell.Biol. *159*, 753-63.

Mahy NL, Perry PE, Gilchrist S, Baldock RA, Bickmore WA. (2002b). Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. J.Cell.Biol. *157*, 579-89.

Maison,C., Bailly,D., Peters,A.H., Quivy,J.P., Roche,D., Taddei,A., Lachner,M., Jenuwein,T., Almouzni,G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat.Genet. *30*, 329-334.

Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ 3rd (1994). Kruppel-associated boxes are potent transcriptional repression domains. Proc.Natl.Acad.Sci.U.S.A. *91*, 4509-4513.

Mark C, Abrink M, Hellman L (1999). Comparative analysis of KRAB zinc finger proteins in rodents and man: evidence for several evolutionarily distinct subfamilies of KRAB zinc finger genes. DNA.Cell.Biol. *18* , 381-396.

Mark,C., Looman,C., Abrink,M., Hellman,L. (2001). Molecular cloning and preliminary functional analysis of two novel human KRAB zinc finger proteins, HKr18 and HKr19. DNA Cell Biol. *20*, 275-286.

Marmorstein R.(2001). Protein modules that manipulate histone tails for chromatin regulation. Nat.Rev.Mol.Cell.Biol. *2*, 422-32.

Marsh,K.L., Dixon,J., Dixon,M.J. (1998). Mutations in the Treacher Collins syndrome gene lead to mislocalization of the nucleolar protein treacle. Hum.Mol.Genet. *7*, 1795-1800.

Martins S, Eikvar S, Furukawa K, Collas P.(2003). HA95 and LAP2 beta mediate a novel chromatin-nuclear envelope interaction implicated in initiation of DNA replication. J.Cell.Biol. *160*, 177-88.

Mateescu,B., England,P., Halgand,F., Yaniv,M., Muchardt,C. (2004). Tethering of HP1 proteins to chromatin is relieved by phosphoacetylation of histone H3. EMBO Rep.

Matsuda E, Agata Y, Sugai M, Katakai T, Gonda H, Shimizu A (2001). Targeting of Kruppel-associated box-containing zinc finger proteins to centromeric heterochromatin. Implication for the gene silencing mechanisms. J.Biol.Chem. *276*, 14222-14229.

Matsui,Y., Zsebo,K., Hogan,B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

McDowell,T.L., Gibbons,R.J., Sutherland,H., O'Rourke,D.M., Bickmore,W.A., Pombo,A., Turley,H., Gatter,K., Picketts,D.J., Buckle,V.J., Chapman,L., Rhodes,D., Higgs,D.R. (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc.Natl.Acad.Sci U.S.A *96*, 13983-13988.

McGhee JD, Nickol JM, Felsenfeld G, Rau DC. (1983). Histone hyperacetylation has little effect on the higher order folding of chromatin. Nucleic.Acids.Res. *11*, 4065-75.

McKittrick E, Gafken PR, Ahmad K, Henikoff S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc.Natl.Acad.Sci.U.S.A. *101*, 1525-30.

Meiboom,M., Murua,E.H., Winkler,S., Nolte,I., Bullerdiek,J. (2004). Molecular characterization and mapping of the canine KRAB zinc finger gene ZNF331. Anim Genet. *35*, 262-263.

Miller,J., McLachlan,A.D., Klug,A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. *4*, 1609-14.

Minc,E., Courvalin,J.C., Buendia,B. (2000). HP1gamma associates with euchromatin and heterochromatin in mammalian nuclei and chromosomes. Cytogenet.Cell Genet. *90*, 279-284.

Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W (1996). Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. Nucleic.Acids.Res. *24* , 4859-4867.

Mounkes L, Kozlov S, Burke B, Stewart CL. (2003). The laminopathies: nuclear structure meets disease. Curr.Opin.Genet.Dev. *13*, 223–230.

Muller,J., Hart,C.M., Francis,N.J., Vargas,M.L., Sengupta,A., Wild,B., Miller,E.L., O'Connor,M.B., Kingston,R.E., Simon,J.A.(2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell. *111*, 197-208.

Muller-Hill,B., Crapo,L., Gilbert,W. (1968). Mutants that make more lac repressor. Proc.Natl.Acad.Sci.U.S.A. *59*, 1259-64.

Nagaki,K., Cheng,Z., Ouyang,S., Talbert,P.B., Kim,M., Jones,K.M., Henikoff,S., Buell,C.R., Jiang,J. (2004). Sequencing of a rice centromere uncovers active genes. Nat.Genet. *36*, 138-145.

Nagaoka M, Nomura W, Shiraishi Y, Sugiura Y (2001). Significant effect of linker sequence on DNA recognition by multi-zinc finger protein. Biochem.Biophys.Res.Commun. *282*, 1001-1007.

Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, Schreiber SL, Evans RM. (1997). Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase.Cell. *89*, 373-80.

Neumann,B., Kubicka,P., and Barow,D.P. (1995). Characteristics of imprinted genes. Nat.Genet. *9*, 12-13.

Ng,H.H. and Bird,A. (1999). DNA methylation and chromatin modification. Curr.Opin.Genet. Dev. *9*:158-163.

Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, Gansmuller A, Chambon P, Losson R (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO.J. *18*, 6385-6395.

Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, Firestein R, Cleary M, Jenuwein T, Herrera RE, Kouzarides T (2001a). Rb targets histone H3 methylation and HP1 to promoters. Nature. *412*, 561-565.

Nielsen,A.L., Oulad-Abdelghani,M., Ortiz,J.A., Remboutsika,E., Chambon,P., Losson,R. (2001b). Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins. Mol.Cell *7*, 729-739.

Nielsen,A.L., Jorgensen,P., Lerouge,T., Cervino,M., Chambon,P., Losson,R. (2004). Nizp1, a Novel Multitype Zinc Finger Protein That Interacts with the NSD1 Histone Lysine Methyltransferase through a Unique C2HR Motif. Mol.Cell Biol. *24*, 5184-5196.

Oh,H.J., Li,Y., Lau,Y.F. (2005). SRY Associates with the Heterochromatin Protein 1 Complex by Interacting with a KRAB Domain Protein. Biol.Reprod. *72*, 407-415.

Okano,M., Xie,S., Li,E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat.Genet. *19*, 219-220.

Okano,M., Bell,D.W., Haber,D.A., Li,E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

O'Neill DW, Schoetz SS, Lopez RA, Castle M, Rabinowitz L, Shor E, Krawchuk D, Goll MG, Renz M, Seelig HP, Han S, Seong RH, Park SD, Agalioti T, Munshi N, Thanos D, Erdjument-Bromage H, Tempst P, Bank A. (2000). An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. Mol.Cell.Biol. *20*, 7572-82.

Osborne,C.S., Chakalova,L., Brown,K.E., Carter,D., Horton,A., Debrand,E., Goyenechea,B., Mitchell,J.A., Lopes,S., Reik,W., Fraser,P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat.Genet. *36*, 1065-71.

Osley,M.A. (2004). H2B ubiquitylation: the end is in sight. Biochim Biophys Acta. *15*, 74-8.

Owen DJ, Ornaghi P, Yang JC, Lowe N, Evans PR, Ballario P, Neuhaus D, Filetici P, Travers AA. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. EMBO J. *19*, 6141-9.

Papworth,M., Moore,M., Isalan,M., Minczuk,M., Choo,Y., Klug,A. (2003). Inhibition of herpes simplex virus 1 gene expression by designer zinc-finger transcription factors. Proc.Natl.Acad.Sci.U.S.A. *100*, 1621-6.

Pal-Bhadra M, Bhadra U, Birchler JA.(2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol.Cell. *9*, 315-27.

Payen E, Verkerk T, Michalovich D, Dreyer SD, Winterpacht A, Lee B, De Zeeuw CI, Grosveld F, Galjart N (1998). The centromeric/nucleolar chromatin protein ZFP-37 may function to specify neuronal nuclear domains. J.Biol.Chem. *273*, 9099-9109.

Peng H, Begg GE, Schultz DC, Friedman JR, Jensen DE, Speicher DW, Rauscher FJ 3rd (2000). Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain- mediated protein-protein interactions. J.Mol.Biol. *295*, 1139-1162.

Peng,H., Feldman,I., Rauscher,F.J., III (2002). Hetero-oligomerization among the TIF family of RBCC/TRIM domain-containing nuclear cofactors: a potential mechanism for regulating the switch between coactivation and corepression. J.Mol.Biol. *320*, 629-644.

Pengue,G., Calabro,V., Bartoli,P.C., Pagliuca,A., Lania,L. (1994). Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. Nucleic Acids Res. *22*, 2908-2914.

Pengue,G., Lania,L. (1996). Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements. Proc.Natl.Acad.Sci.U.S.A *93*, 1015-1020.

Peters,A.H., O'Carroll,D., Scherthan,H., Mechtler,K., Sauer,S., Schofer,C., Weipoltshammer,K., Pagani,M., Lachner,M., Kohlmaier,A., Opravil,S., Doyle,M., Sibilia,M., Jenuwein,T. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

Peters,A.H., Mermoud,J.E., O'Carroll,D., Pagani,M., Schweizer,D., Brockdorff,N., Jenuwein,T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat.Genet. *30*, 77-80.

Peters,A.H., Kubicek,S., Mechtler,K., O'Sullivan,R.J., Derijck,A.A., Perez-Burgos,L., Kohlmaier,A., Opravil,S., Tachibana,M., Shinkai,Y., Martens,J.H., Jenuwein,T. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol.Cell *12*, 1577-1589.

Peterson CL. (2002). HDAC's at work: everyone doing their part. Mol Cell. *9*, 921-2.

Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ. (1996). ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.Hum.Mol.Genet. *5*, 1899-907.

Platero,J.S., Csink,A.K., Quintanilla,A., Henikoff,S. (1998). Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in Drosophila. J.Cell Biol. *140*, 1297-1306.

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B.(2002). Xist RNA and the mechanism of X chromosome inactivation. Annu.Rev.Genet. *36*, 233-78.

Poot,R.A., Bozhenok,L., van den Berg,D.L., Steffensen,S., Ferreira,F., Grimaldi,M., Gilbert,N., Ferreira,J., Varga-Weisz,P.D. (2004). The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. Nat.Cell.Biol. *6*, 1236-44.

Ptashne,M. Specific binding of the lambda phage repressor to lambda DNA. (1967). Nature *214*, 232-4.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc.Natl.Acad.Sci.U.S.A. *97*, 5237-5242.

Ransom,D.G., Bahary,N., Niss,K., Traver,D., Burns,C., Trede,N.S., Paffett-Lugassy,N., Saganic,W.J., Lim,C.A., Hersey,C., Zhou,Y., Barut,B.A., Lin,S., Kingsley,P.D., Palis,J., Orkin,S.H., Zon,L.I. (2004). The Zebrafish moonshine Gene Encodes Transcriptional Intermediary Factor 1gamma, an Essential Regulator of Hematopoiesis. PLoS.Biol. *2*, E237.

Ravasi,T., Huber,T., Zavolan,M., Forrest,A., Gaasterland,T., Grimmond,S., Hume,D.A. (2003). Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Res. *13*, 1430-1442.

Rayasam,G.V., Wendling,O., Angrand,P.O., Mark,M., Niederreither,K., Song,L., Lerouge,T., Hager,G.L., Chambon,P., Losson,R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. *22*, 3153-3163.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature. *406*, 593-599.

Reed,K.C., Mann,D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. *13*, 7207-7221.

Remboutsika,E., Lutz,Y., Gansmuller,A., Vonesch,J.L., Losson,R., Chambon,P. (1999). The putative nuclear receptor mediator TIF1alpha is tightly associated with euchromatin. J.Cell Sci. *112 (Pt 11)*, 1671-1683.

Resch,A., Xing,Y., Modrek,B., Gorlick,M., Riley,R., Lee,C. (2004). Assessing the impact of alternative splicing on domain interactions in the human proteome. J.Proteome.Res. *3*, 76-83.

Rhoades, R. and Pflanzer, R. Human Physiology [3rd Edition]. (1996). Saunders College Publishing.

Rice,J.C., Allis,C.D. (2001). Histone methylation versus histone acetylation: new insights into epigenetic regulation. Curr.Opin.Cell Biol. *13*, 263-273.

Rice,J.C., Briggs,S.D., Ueberheide,B., Barber,C.M., Shabanowitz,J., Hunt,D.F., Shinkai,Y., Allis,C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. Mol.Cell *12*, 1591-1598.

Richards EJ, Elgin SC (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell. *108*, 489-500.

Robzyk,K., Recht,J., Osley,M.A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science. *287*, 501-4.

Romano G.Histone variants during sea urchin development. (1992). Cell.Biol.Int.Rep.*16*,197-206.

Rooney,J.W., Calame,K.L. (2001). TIF1beta functions as a coactivator for C/EBPbeta and is required for induced differentiation in the myelomonocytic cell line U937. Genes Dev. *15*, 3023-3038.

Rosenberg,U.B., Schroeder,C., Preiss,A., Keinlin,A., Cote,S., Riede,I., and Jackle,H. (1986). Structural homology of the product of the *Drosophila Krüppel* gene with *Xenopus* transcription factor IIIA. Nature. *319*, 336-339.

Rue SW, Kim BW, Jun DY, Kim YH (2001). Nucleotide sequence and cell cycle-associated differential expression of ZF5128, a novel Kruppel type zinc finger protein gene. Biochim.Biophys.Acta. *1522*, 230-237.

Ryan RF, Schultz DC, Ayyanathan K, Singh PB, Friedman JR, Fredericks WJ, Rauscher FJ 3rd (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and

euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol.Cell.Biol. *19*, 4366-4378.

Sabater L, Ashhab Y, Caro P, Kolkowski EC, Pujol-Borrell R, Dominguez O (2002). Identification of a KRAB-containing zinc finger protein, ZNF304, by AU- motif-directed display method and initial characterization in lymphocyte activation. Biochem.Biophys.Res.Commun. *293*, 1066-1072.

Sabbattini,P., Lundgren,M., Georgiou,A., Chow,C., Warnes,G., Dillon,N. (2001). Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. EMBO J. *20*, 2812-2822.

Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, Turner BM, Zink D.(1999). Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J.Cell.Biol. *146*, 1211-26.

Sage,B.T., Jones,J.L., Holmes,A.L., Wu,M.D., Csink,A.K. (2005). Sequence Elements in cis Influence Heterochromatic Silencing in trans. Mol.Cell Biol. *25*, 377-388.

Sambrook, J. and and Russell, D. W. Molecular Cloning. Irwin, N. and Janssen K. A laboratory manual. [3rd Edition]. 2001. Cold Spring Harbor Laboratory Press.

Santisteban MS, Kalashnikova T, Smith MM. (2000). Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes.Cell. 103, 411-22.

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature. *419*, 407-11.

Sarg B, Helliger W, Talasz H, Koutzamani E, Lindner HH. (2004). Histone H4 hyperacetylation precludes histone H4 lysine 20 trimethylation. J.Biol.Chem.*279*, 53458-64.

Sarraf,S.A., Stancheva,I. (2004). Methyl-CpG Binding Protein MBD1 Couples Histone H3 Methylation at Lysine 9 by SETDB1 to DNA Replication and Chromatin Assembly. Mol.Cell *15*, 595-605.

Saunders WS, Chue C, Goebl M, Craig C, Clark RF, Powers JA, Eissenberg JC, Elgin SC, Rothfield NF, Earnshaw WC. (1993). Molecular cloning of a human homologue of Drosophila heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity. J.Cell.Sci. *104*, 573-82.

Saurin,A.J., Borden,K.L., Boddy,M.N., Freemont,P.S. (1996). Does this have a familiar RING? Trends Biochem Sci. *21*, 208-14.

Schaufele F, Enwright JF 3rd, Wang X, Teoh C, Srihari R, Erickson R, MacDougald OA, Day RN. (2001). CCAAT/enhancer binding protein alpha assembles essential cooperating factors in common subnuclear domains. Mol Endocrinol. *15*, 1665-76.

Schotta,G., Ebert,A., Krauss,V., Fischer,A., Hoffmann,J., Rea,S., Jenuwein,T., Dorn,R., Reuter,G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. *21*, 1121-1131.

Schotta,G., Lachner,M., Sarma,K., Ebert,A., Sengupta,R., Reuter,G., Reinberg,D., Jenuwein,T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. *18*, 1251-1262.

Schultz DC, Friedman JR, Rauscher FJ 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes.Dev. *15*, 428-443.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes.Dev. *16*, 919-932.

Seeler,J.S., Marchio,A., Sitterlin,D., Transy,C., Dejean,A. (1998). Interaction of SP100 with HP1 proteins: a link between the promyelocytic leukemia-associated nuclear bodies and the chromatin compartment. Proc.Natl.Acad.Sci.U.S.A *95*, 7316-7321.

Seeler,J.S., Marchio,A., Losson,R., Desterro,J.M., Hay,R.T., Chambon,P., Dejean,A. (2001). Common properties of nuclear body protein SP100 and TIF1alpha chromatin factor: role of SUMO modification. Mol.Cell Biol. *21*, 3314-3324.

Segal,D.J. (2002). The use of zinc finger peptides to study the role of specific factor binding sites in the chromatin environment. Methods *26*, 76-83.

Shannon M, Kim J, Ashworth L, Branscomb E, Stubbs L (1998). Tandem zinc-finger gene families in mammals: insights and unanswered questions. DNA.Seq. *8*, 303-315.

Shannon,M., Stubbs,L. (1999). Molecular characterization of Zfp54, a zinc-finger-containing gene that is deleted in the embryonic lethal mutation tw18. Mamm.Genome *10*, 739-743.

Shannon,M., Hamilton,A.T., Gordon,L., Branscomb,E., Stubbs,L. (2003). Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. Genome Res. *13*, 1097-1110.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell. *119*, 941-53.

Shopland LS, Johnson CV, Byron M, McNeil J, Lawrence JB. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. J.Cell.Biol. *162*, 981-90.

Shukunami,C., Shigeno,C., Atsumi,T., Ishizeki,K., Suzuki,F., Hiraki,Y. (1996). Chondrogenic differentiation of clonal mouse embryonic cell line ATDC5 in vitro: differentiation-dependent gene expression of parathyroid hormone (PTH)/PTH-related peptide receptor. J.Cell Biol. *133*, 457-468.

Singh,P.B., Georgatos,S.D. (2002). HP1: facts, open questions, and speculation. J.Struct.Biol. *140*, 10-16.

Skapek SX, Jansen D, Wei TF, McDermott T, Huang W, Olson EN, Lee EY (2000). Cloning and characterization of a novel Kruppel-associated box family transcriptional repressor that interacts with the retinoblastoma gene product, RB. J.Biol.Chem. *275*, 7212-7223.

Sleutels F, Zwart R, Barlow DP.(2002). The non-coding Air RNA is required for silencing autosomal imprinted genes.Nature. *415*, 810-3.

Solter D, Knowles BB (1978). Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). Proc.Natl.Acad.Sci.U.S.A. *75*, 5565-5569.

Spector DL (2001). Nuclear domains. J.Cell.Sci. *114*, 2891-2893.

Stern M, Jensen R, Herskowitz I. (1984). Five SWI genes are required for expression of the HO gene in yeast. J.Mol.Biol. *178*, 853-68.

Stone,S.L., Hauksdottir,H., Troy,A., Herschleb,J., Kraft,E., Callis,J. (2005). Functional Analysis of the RING-Type Ubiquitin Ligase Family of Arabidopsis. Plant Physiol. *137*, 13-30.

Stopka,T., Skoultchi,A.I. (2003). The ISWI ATPase Snf2h is required for early mouse development. Proc.Natl.Acad.Sci.U.S.A. *100*, 14097-102.

Strahl,B.D., Allis,C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Strickland S, Mahdavi V (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell. *15*, 393-403.

Stuurman,N., Heins,S., Aebi,U. (1998). Nuclear lamins: their structure, assembly, and interactions. J Struct Biol. *122*, 42-66.

Sudarsanam,P., Winston,F. (2000). The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. Trends Genet. *16*, 345-351.

Sullivan,B., Karpen,G. (2001). Centromere identity in Drosophila is not determined in vivo by replication timing. J.Cell Biol. *154*, 683-690.

Sun,Y., Gou,D.M., Liu,H., Peng,X., Li,W.X. (2003). The KRAB domain of zinc finger gene ZNF268: a potential transcriptional repressor. IUBMB.Life *55*, 127-131.

Sutherland HG, Mumford GK, Newton K, Ford LV, Farrall R, Dellaire G, Caceres JF, Bickmore WA (2001). Large-scale identification of mammalian proteins localized to nuclear sub-compartments. Hum.Mol.Genet. *10*, 1995-2011.

Sutherland,H.G., Lam,Y.W., Briers,S., Lamond,A.I., Bickmore,W.A. (2004). 3D3/lyric: a novel transmembrane protein of the endoplasmic reticulum and nuclear envelope, which is also present in the nucleolus. Exp.Cell Res. *294*, 94-105.

Syntichaki,P., Topalidou,I., Thireos,G. (2000). The Gcn5 bromodomain co-ordinates nucleosome remodelling. Nature *404*, 414-417.

Szentirmay MN, Sawadogo M. (2000). Spatial organization of RNA polymerase II transcription in the nucleus. Nucleic.Acids.Res. *28*, 2019-25.

Tachibana,M., Sugimoto,K., Fukushima,T., Shinkai,Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J.Biol.Chem. *276*, 25309-25317.

Taddei,A., Maison,C., Roche,D., Almouzni,G. (2001). Reversible disruption of pericentric heterochromatin and centromere function by inhibiting deacetylases. Nat.Cell Biol. *3*, 114-120.

Tamaru,H., Selker,E.U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature *414*, 277-283.

Tan,W., Zheng,L., Lee,W.H., Boyer,T.G. (2004a). Functional dissection of transcription factor ZBRK1 reveals zinc fingers with dual roles in DNA-binding and BRCA1-dependent transcriptional repression. J.Biol.Chem. *279*, 6576-6587.

Tan,W., Kim,S., Boyer,T.G. (2004b). Tetrameric oligomerization mediates transcriptional repression by the BRCA1-dependent KRAB-zinc finger protein ZBRK1. J.Biol.Chem.

Tanaka K, Tsumaki N, Kozak CA, Matsumoto Y, Nakatani F, Iwamoto Y, Yamada Y (2002). A Kruppel-associated box-zinc finger protein, NT2, represses cell-type- specific promoter activity of the alpha 2(XI) collagen gene. Mol.Cell.Biol. *22*, 4256-4267.

Tariq,M., Saze,H., Probst,A.V., Lichota,J., Habu,Y., Paszkowski,J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc.Natl.Acad.Sci.U.S.A *100* , 8823-8827.

Tate,P., Lee,M., Tweedie,S., Skarnes,W.C., Bickmore,W.A. (1998). Capturing novel mouse genes encoding chromosomal and other nuclear proteins. J.Cell Sci. *111*, 2575-2585.

Taverna SD, Coyne RS, Allis CD. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell. *110*, 701-11.

Theunissen,O., Rudt,F., Guddat,U., Mentzel,H., Pieler,T. (1992). RNA and DNA binding zinc fingers in Xenopus TFIIIA. Cell. *71*, 679-90.

Tolhuis,B., Palstra,R.J., Splinter,E., Grosveld,F., de Laat,W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol.Cell. *10*, 1453-65.

Tommerup,N., Vissing,H. (1995). Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. Genomics *27*. 259-264. Ref ID: 319

Trockenbacher,A., Suckow,V., Foerster,J., Winter,J., Krauss,S., Ropers,H.H., Schneider,R., Schweiger,S. (2001). MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. Nat Genet. *29*, 287-94.

Tsukiyama,T. (2002). The in vivo functions of ATP-dependent chromatin-remodelling factors. Nat Rev.Mol.Cell.Biol. *3*, 422-9.

Tullis,K.M., Krebs,C.J., Leung,J.Y., Robins,D.M. (2003). The regulator of sex-limitation gene, rsl, enforces male-specific liver gene expression by negative regulation. Endocrinology. *144*, 1854-1860.

Tunnacliffe,A., Liu,L., Moore,J.K., Leversha,M.A., Jackson,M.S., Papi,L., Ferguson-Smith,M.A., Thiesen,H.J., Ponder,B.A. (1993). Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. Nucleic Acids Res. *21*, 1409-1417.

Underhill C, Qutob MS, Yee SP, Torchia J (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. J.Biol.Chem. *275*, 40463-40470.

Urrutia,R. (2003). KRAB-containing zinc-finger repressor proteins. Genome Biol. *4*, 231.

van Leeuwen F, Gafken PR, Gottschling DE. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell. *109*, 745-56.

Venturini,L., You,J., Stadler,M., Galien,R., Lallemand,V., Koken,M.H., Mattei,M.G., Ganser,A., Chambon,P., Losson,R., de The,H. (1999). TIF1gamma, a novel member of the transcriptional intermediary factor 1 family. Oncogene *18*, 1209-1217.

Verdel,A., Jia,S., Gerber,S., Sugiyama,T., Gygi,S., Grewal,S.I., Moazed,D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Verschure PJ, van Der Kraan I, Manders EM, van Driel R (1999). Spatial relationship between transcription sites and chromosome territories. J.Cell.Biol. *147*, 13-24.

Villa,A., Strina,D., Frattini,A., Faranda,S., Macchi,P., Finelli,P., Bozzi,F., Susani,L., Archidiacono,N., Rocchi,M., Vezzoni,P. (1996). The ZNF75 zinc finger gene subfamily: isolation and mapping of the four members in humans and great apes. Genomics *35*, 312-320.

Vissing H, Meyer WK, Aagaard L, Tommerup N, Thiesen HJ (1995). Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. FEBS.Lett. *369*, 153-157.

Volpe,T.A., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I., Martienssen,R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Volpi EV, Chevret E, Jones T, Vatcheva R, Williamson J, Beck S, Campbell RD, Goldsworthy M, Powis SH, Ragoussis J, Trowsdale J, Sheer D (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. J.Cell.Sci. 113, 1565-1576.

Voncken,J.W., Roelen,B.A., Roefs,M., de Vries,S., Verhoeven,E., Marino,S., Deschamps,J., van Lohuizen,M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition.
Proc.Natl.Acad.Sci.U.S.A. *100*, 2468-73.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat.Genet. *23*, 62-6.

Wagner,S., Hess,M.A., Ormonde-Hanson,P., Malandro,J., Hu,H., Chen,M., Kehrer,R., Frodsham,M., Schumacher,C., Beluch,M., Honer,C., Skolnick,M., Ballinger,D., Bowen,B.R. (2000). A broad role for the zinc finger protein ZNF202 in human lipid metabolism. J.Biol.Chem. *275*, 15685-15690.

Wakimoto,B.T. (1998). Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. Cell *93*, 321-324.

Wang,H., An,W., Cao,R., Xia,L., Erdjument-Bromage,H., Chatton,B., Tempst,P., Roeder,R.G., Zhang,Y. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol.Cell *12*, 475-487.

Wang J, Shiels C, Sasieni P, Wu PJ, Islam SA, Freemont PS, Sheer D.(2004). Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. J.Cell.Biol. *164*, 515-26.

Weber,P., Cammas,F., Gerard,C., Metzger,D., Chambon,P., Losson,R., Mark,M. (2002). Germ cell expression of the transcriptional co-repressor TIF1beta is required for the maintenance of spermatogenesis in the mouse. Development *129*, 2329-2337.

Whitehouse I, Flaus A, Cairns BR, White MF, Workman JL, Owen-Hughes T. (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. Nature. *400*, 784-7.

Wick,M.J., Ann,D.K., Lee,N.M., Loh,H.H. (1995). Isolation of a cDNA encoding a novel zinc-finger protein from neuroblastoma x glioma NG108-15 cells. Gene *152*, 227-232.

Widom J. (1998). Structure, dynamics, and function of chromatin in vitro. Annu Rev Biophys Biomol Struct. *27*, 285-327.

Williams,A.J., Khachigian,L.M., Shows,T., Collins,T. (1995). Isolation and characterization of a novel zinc-finger protein with transcription repressor activity. J.Biol.Chem. *270*, 22143-52.

Williams RR, Broad S, Sheer D, Ragoussis J (2002). Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. Exp.Cell.Res. *272*, 163-175.

Witzgall,R., O'Leary,E., Leaf,A., Onaldi,D., Bonventre,J.V. (1994). The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. Proc.Natl.Acad.Sci.U.S.A.*91*, 4514-4518.

Wolffe AP, Hayes JJ. (1999). Chromatin disruption and modification.Nucleic Acids Res. *27*, 711-20.

Wolffe,A.P. Centromeric chromatin. Histone deviants. (1995). *5*, 452-4.

Xing,Y., Johnson,C.V., Moen,P.T.Jr., McNeil,J.A., Lawrence,J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. J.Cell.Biol. *131*,1635-47.

Xu,G.L., Bestor,T.H., Bourc'his,D., Hsieh,C.L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,J.J., Viegas-Pequignot,E.(1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. *402*, 187-91.

Yang,L., Xia,L., Wu,D.Y., Wang,H., Chansky,H.A., Schubach,W.H., Hickstein,D.D., Zhang,Y. (2002). Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor. Oncogene *21*, 148-152.

Yano,K., Ueki,N., Oda,T., Seki,N., Masuho,Y., Muramatsu,M. (2000). Identification and characterization of human ZNF274 cDNA, which encodes a novel kruppel-type zinc-finger protein having nucleolar targeting ability. Genomics. *65*, 75-80.

Ye,Q., Worman,H.J. (1996). Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. J.Biol.Chem. *271*, 14653-14656.

Zaidi,S.K., Sullivan,A.J., van Wijnen,A.J., Stein,J.L., Stein,G.S., Lian,J.B. (2002). Integration of Runx and Smad regulatory signals at transcriptionally active subnuclear sites. Proc.Natl.Acad.Sci U S A. *99*, 8048-53.

Zeng,C., Kim,E., Warren,S.L., Berget,S.M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. EMBO J. *16*, 1401-12.

Zhang,C.L., McKinsey,T.A., Olson,E.N. (2002). Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. Mol.Cell Biol. *22*, 7302-7312.

Zhang,W.H., Srihari,R., Day,R.N., Schaufele,F. (2001). CCAAT/enhancer-binding protein alpha alters histone H3 acetylation at large subnuclear domains. J Biol Chem. *276*, 40373-6.

Zhang,Y., LeRoy,G., Seelig,H.P., Lane,W.S., Reinberg,D.(1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell. *95*, 279-89.

Zhang,Y., Ng,H.H., Erdjument-Bromage,H., Tempst,P., Bird,A., Reinberg,D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. *13*, 1924-35.

Zheng L, Pan H, Li S, Flesken-Nikitin A, Chen PL, Boyer TG, Lee WH (2000). Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1. Mol.Cell. *6*, 757-768.

Zheng P, Guo Y, Niu Q, Levy DE, Dyck JA, Lu S, Sheiman LA, Liu Y. (1998). Proto-oncogene PML controls genes devoted to MHC class I antigen presentation. Nature. *396*, 373-6.

Zhou L, Zhu C, Luo K, Li Y, Pi H, Yuan W, Wang Y, Huang C, Liu M, Wu X (2002). Identification and characterization of two novel zinc finger genes, ZNF359 and ZFP28, in human development. Biochem.Biophys.Res.Commun. *295*, 862-868.

Zlatanova,J., Leuba,S.H., van Holde,K. (1999). Chromatin structure revisited. Crit Rev Eukaryot Gene Expr. *9*, 245-55.

Zilberman,D., Cao, X., Jacobsen, S.E.(2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science. *299*, 716-9.

Zink,D., Amaral,M.D., Englmann,A., Lang,S., Clarke,L.A., Rudolph,C., Alt,F., Luther,K., Braz,C., Sadoni,N., Rosenecker,J., Schindelhauer,D. (2004). Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. J.Cell Biol. *166*, 815-825.